ESTIMATING V&V RESOURCE REQUIREMENTS AND SCHEDULE IMPACT

Michelle L. Kilikauskas
Dirk Brade
Robert M. Gravitz
David H. Hall
Martha L. Hoppus
Ronald L. Ketcham
Robert O. Lewis
Michael L. Metz

An Invited Paper for Session B4 of the

Foundations for V&V in the 21st Century Workshop

(Foundations '02)

Held at the Johns Hopkins University Applied Physics Laboratory
Laurel, Maryland (USA)
October 22-24, 2002

Introduction

This paper addresses methods used in the community to predict the level of verification and validation (V&V) resources required for various levels of model and simulation (M&S) credibility. We discuss several approaches to predicting V&V resource requirements. In addition, we present a number of case studies based on the experiences of programs which needed to accredit M&S for their use; these case studies provide background on the cost and schedule impact of V&V activities and accreditation requirements. Finally, we draw some conclusions about the ability of the community to predict and plan for V&V resource requirements, and we make some recommendations for future research in this area.

Perceptions of V&V Resource Requirements

Some common misperceptions of V&V resource requirements are directly related to common misperceptions of VV&A. One of these is the idea that: "VV&A" IS A ONE-TIME EVENT. This is clearly not true, since accreditation is needed for each new use of the M&S, and validation in particular is done (or should be done) in the context of an intended use. Thus V&V is never "completed" for all time and uses. This has implications for an accreditation authority or a model proponent's resource plans for V&V activities, since they may be thinking that because someone else already accredited the M&S, they don't have to do any more work.

Another common misperception is that: VV&A IS A "CHECK IN THE BOX". Even though VV&A is required by policy, the reason we conduct V&V activities is that they are essential to reduce the risk of inappropriate decisions based on erroneous model outputs. In this context V&V activities constitute risk reduction efforts for the accrediting authority's use of M&S results to influence their decisions. The "check in the box" mentality leads some users to grossly underestimate the requirements for V&V resources and the schedule impact of those activities.

One very common misperception is that: "SOFTWARE V&V = SIMULATION V&V". This is not true. The Carnegie Mellon University Software Engineering Institute (SEI) defines <u>software verification</u> as "the process of determining whether or not the products of a given phase of a software development process fulfill the requirements established…"[1] and <u>software validation</u> as "the process of evaluating software at the end of its development to insure that it is free from failures and complies with its requirements."[2] Law and Kelton describe software verification as the process of determining whether the product meets the customer's specifications and software validation as the process of determining whether the product meets the customer's needs.[3] However, for the special case of models and simulations, the official Department of Defense (DOD) dictionary defines <u>simulation verification</u> as " the process of determining that a model or simulation implementation accurately represents the developer's conceptual description and specifications", and <u>simulation validation</u> as the process of determining the degree to which a

[1] *Introduction to Software Verification and Validation*, SEI Curriculum Module SEI-CM-13-1.1, James S. Collofello, Carnegie Mellon University Software Engineering Institute, December 1988, pg 4

[2] *ibid*

[3] *Simulation Modeling and Analysis (Third Edition)*, Averill M. Law and W. David Kelton, McGraw-Hill, 2000

model or simulation is an accurate representation of the real world from the perspective of the intended uses of the model or simulation"[4] (accreditation is a decision by the user as to whether a model or simulation meets that user's needs for a particular application). So by these definitions, *software V&V* activities are included in *simulation verification.* On the other hand, comparing simulation predictions with actual field test data (or simulator data, or benchmarking, etc.) is not a part of normal software V&V activities: *simulation validation* is not really a software issue per se. Simulation validation can be very expensive, but it is not part of a software V&V plan developed during software development. Thus this misconception can cause a program to grossly underestimate V&V costs, if they think that by doing software validation they are accomplishing simulation validation.

Quite a few people think that: VV&A IS A "SOFTWARE DEVELOPMENT ISSUE". While it is true that software verification and validation are normally done during software development as a means of catching errors, they are by no means a sufficient basis for accreditation. Simulation validation and accreditation are really User concerns that require resources beyond those expended during software development. VV&A cannot be contracted out to a software IV&V group, because they only address the part of overall M&S credibility having to do with the accuracy of the software and the conceptual model, not actual correlation with test results (results validation) or expert opinion (face validation).

A related misconception can be expressed as: "VV&A IS ONE WORD, AND V+V=A". By this we mean that many people feel that "VV&A" is all part of one set of activities, taken care of by someone else, usually the model developer. But accreditation is a completely different activity from V&V and is normally a User function (as is simulation validation), whereas software V&V is primarily a Developer function. The information generated through V&V activities is used in the accreditation assessment, but a user will also require other information in addition to V&V results to make an accreditation decision, such as M&S functionality and usability. Thus V+V does not equal A. And of course this means that reaching an accreditation decision will require resources beyond those budgeted for traditional software V&V activities.

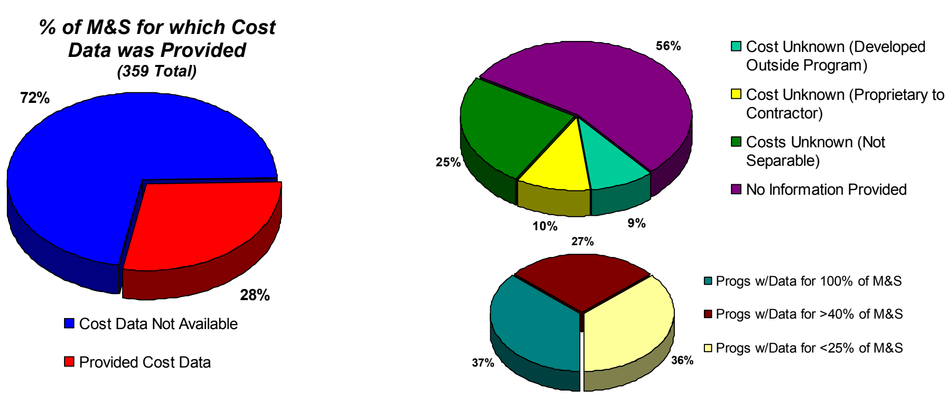
Motivation for V&V cost estimation

A last, but certainly not least perception that many people have (Program Managers in particular) is that V&V: "COSTS TOO MUCH, TAKES TOO LONG". This is really the primary motivation for developing a cost estimation process for V&V activities. If we can show what V&V actually takes in terms of cost, manpower, schedule, etc., we can help to dispel the fairly prevalent attitude of Department of Defense Program Managers that "it's too expensive to do V&V, but policy says we have to do it, so we'll pay lip service to it". We will see that V&V is too expensive only if V&V activities are not focused on supporting an accreditation decision. We will see in the discussion below that the resources required for V&V activities can be minimized by focusing on accreditation requirements and on reducing the risk of using M&S results during the development of the V&V plan.

[4] *DoD Dictionary of Military and Associated Terms*, Joint Publication 1-02, Department of Defense, 19 December 2001

There are a number of motivations for programs to make use of models and simulations as part of their process of developing systems, not least of which is a cost-benefit argument. Whether the use of M&S reduces acquisition program costs is a somewhat contentious issue, but it is universally agreed that the use of M&S provides some cost benefit. In 1995 The Target Interaction, Lethality and Vulnerability (TILV) Science and Technology (S&T) coordination group made an attempt to estimate the savings experienced by acquisition programs in their use of M&S: their estimated return on investment (ROI) varied between 15 and 40%[5]. This was for a number of defense acquisition programs, varying from aircraft to weapons systems in all the Services. However, those numbers were very rough estimates only, since detailed M&S cost data were not available at the time.

A later study conducted by Hicks & Associates for the Director, Operational Test and Evaluation (DOT&E) focused on the use of M&S in acquisition programs, and their costs to those programs[6]. They evaluated M&S use and associated costs across 22 acquisition programs, for a total of 359 M&S used by those programs. The primary conclusion of the study with respect to cost was that M&S development and application cost data are not readily available within acquisition programs. The figure below shows the available cost data (this is for M&S use as a whole, not for VV&A costs). As the figure shows, only 28% of the programs could provide any cost data at all, and only 37% of those could produce 100% of the data on what the M&S had cost. Now it is entirely possible that the approximately half of the programs who did not provide any information simply found it too much trouble to figure out their costs just to respond to a survey. However, it is telling that of those who did respond, most did not have anything close to complete information.



DOT&E Survey M&S Cost Data

[5] *DoD Master Plan for Target Interaction, Lethality and Vulnerability (TILV) Science and Technology (S&T) Programs, Volume I: Classical Ballistic Threats*, Department of Defense, 4 May 1995 revision.

[6] *The Use of Modeling and Simulation (M&S) Tools in Acquisition Program Offices: Results of a Survey*, Anne Hillegas, John Backschies, Michael Donley, R.Cliff Duncan, William Edgar, Hicks & Associates, 31 Jan 2001

Most of the programs who responded simply did not track M&S costs, let alone V&V costs. Only one of the programs surveyed had a detailed Work Breakdown Structure (WBS) for M&S tasking (including VV&A). The study concluded that there is a lack of management visibility into program expenditures for M&S activities in general, in part because standard cost accounting procedures do not provide for segregation, reporting or tracking of M&S costs. Another complicating factor is that M&S activities often are not listed as deliverable items in contracts, meaning that the contractor is not under any obligation to report the expenses associated with M&S activities even if they could do so. And programs themselves are not required to track M&S expenditures, so they don't track them. (This is generally the case when the contract was for development of some item like an airplane or missile and development of M&S is just a subtask involved in making the primary deliverable - the hardware. Contracts that are let specifically for M&S development may be more likely to track V&V costs.) All of this exacerbates the problem of coming up with a good way to predict V&V costs. If we don't even know what the M&S cost, how are we to predict the resource requirements for V&V?

A recent major study by the National Institute of Standards (NIST) estimated the huge industry-wide cost of inadequate software testing (in terms of significant percentages of the Gross National Product). The NIST study also pointed out that the emphasis in software development has shifted over the years in terms of the percentage of time spent in software testing (which is part of V&V) vice other activities. That study pointed out that, "Historically, software development focused on writing code and testing specific lines of that code. Very little effort was spent on determining its fit within a larger system. Testing was seen as a necessary evil to prove to the consumer that the product worked."[7] The table below is reproduced from that NIST report, showing the allocation of labor effort across the various phases of software development projects by decade. [8]

<table>
<tr><th></th><th>Requirements Analysis</th><th>Preliminary Design</th><th>Detailed Design</th><th>Coding and Unit Testing</th><th>Integration and Test</th><th>System Test</th></tr>
<tr><td>1960s – 1970s</td><td colspan="3">10%</td><td>80%</td><td colspan="2">10%</td></tr>
<tr><td>1980s</td><td colspan="2">20%</td><td colspan="2">60%</td><td colspan="2">20%</td></tr>
<tr><td>1990s</td><td>40%</td><td colspan="2">30%</td><td colspan="3">30%</td></tr>
</table>

Allocation of Effort

As can be seen from the table, since those early phases of software development, more and more thought has gone into developing requirements and design prior to coding, and fewer resources (as a percentage of labor hours) have gone into testing. While this may be a result of better

[7] *The Economic Impacts of Inadequate Infrastructure for Software Testing*, National Institute of Standards Acquisition and Assistance Division, Planning Report 02-3, May 2002, pp ES-4 – ES-6

[8] *Ibid,* pg ES-5; although produced on page ES-5 of the NIST Planning Report, the information in the table was obtained from *Formalizing Use Cases with Message Sequence Charts,* Andersson, M. and J. Bergstrand, 1995, Unpublished Master's thesis. Lund Institute of Technology, Lund, Sweden planning, it makes it difficult to estimate V&V resource requirements as a percentage of overall software development cost. The NIST study concluded that, "Testing activities are conducted throughout all the development phases shown in (the table). Formal testing conducted by independent test groups accounts for about 20 percent of labor costs. However, estimates of total labor resources spent testing by all parties range from 30 to 90 percent."[9] A wide variation is apparent in these estimates.

Approach taken in this paper

In order to address the resource requirements for VV&A, and the difficulty of estimating them, we will start by identifying a number of factors that influence the scope of required V&V activities, including the complexity both of the M&S and the application. It may be possible to conceive of a comprehensive, exhaustive approach to verification, and even of an extensive approach to validation limited only by the amount of reliable validation data available. However, most applied VV&A approaches in the DOD community all take into account that the risk associated with the application greatly influences the scope of V&V required. The higher the risk associated with the application of M&S to support decision making means the more evidence required of M&S credibility. So risk and the availability of information about the M&S and of data to support validation have a direct influence on the resource requirements for V&V.

Next we will present the "state of the art" of V&V resource estimation, from several perspectives. The first is an examination of "Risk-Based" approaches both in the United States and in the United Kingdom. A managed investment approach for VV&A will be discussed. Finally, a V&V Cost Estimating Tool (CET) will be described.

Following the state of the art discussion, a number of case histories will describe the approaches and cost experiences of several programs, including missile systems programs, aircraft programs, a major training system development, and a classic VV&A process development program that tracked the cost of V&V activities over a number of M&S via a Work Breakdown Structure (WBS).

Last we present an analysis of the state of the art and of the case histories, to describe the relative merits of the various approaches. A set of recommendations are put forward that will enhance the state of the art for estimating resource requirements for V&V activities.

Primary Pertinent Literature

The following were the **primary pertinent literature** and resources used in developing this paper:

---

[9] *Ibid*, pg ES-5

The Application of VV&A in Promoting the Credible Employment of M&S within the Joint Strike Fighter Program, Ronald L. Ketcham and Maj. Steven Bishop, Proceedings of the 2002 European Simulation MultiConference, Society for Computer Simulation Europe, June 2002

How to VV&A Without Really Trying: SMART Lessons Learned, David H. Hall, Michelle L. Kilikauskas, Dennis K. Laack, Dr. Paul R. Muessig, Barry O'Neal, Chester Richardson, Karl Simecka, Willie Stewart, Dr. Stewart W. Turner, Joint Technical Coordinating Group on Aircraft Survivability, JTCG/AS-07-M-009, November 1997

Crosstalk, June 2002

An Integrated Approach to Evaluating Simulation Credibility, Dr. Paul R. Muessig, Dennis R. Laack; John J. Wrobleski, Proceedings of the 2000 Summer Computer Simulation Conference, Society for Computer Simulation, 2000

Introduction to Software Verification and Validation, SEI Curriculum Module SEI-CM-13-1.1, James S. Collofello, Carnegie Mellon University Software Engineering Institute, December 1988
Simulation Modeling and Analysis (Third Edition), Averill M. Law and W. David Kelton, McGraw-Hill, 2000

DoD Dictionary of Military and Associated Terms, Joint Publication 1-02, Department of Defense, 19 December 2001

DoD Master Plan for Target Interaction, Lethality and Vulnerability (TILV) Science and Technology (S&T) Programs, Volume I: Classical Ballistic Threats, Department of Defense, 4 May 1995 revision.

The Use of Modeling and Simulation (M&S) Tools in Acquisition Program Offices: Results of a Survey, Anne Hillegas, John Backschies, Michael Donley, R.Cliff Duncan, William Edgar, Hicks & Associates, 31 Jan 2001

The Economic Impacts of Inadequate Infrastructure for Software Testing, National Institute of Standards Acquisition and Assistance Division, Planning Report 02-3, May 2002

Department of Defense Modeling and Simulation Management, Department of Defense Directive (DODD) 5000.59

Managing Technology, The Strategic View, Steele, Lowell W. 1989. McGraw-Hill

Optimizing the Selection of VV&A Activities A Risk/Benefit Approach, Muessig, P. R., Laack, D. R. and Wrobleski, J. J., Proceedings of the 1997 Summer Computer Simulation Conference Arlington VA. pp 855-860

Verification, Validation and Accreditation of Models and Simulations Used for Test and Evaluation: A Risk/Benefit Based Approach, Chris Mugridge, Defence Evaluation and Research Agency, UK, March 1999

Cost vs. Credibility: How Much V&V Is Enough?, Muessig, P.R., Society for Computer Simulation

VV&A from A to Z, Dr. Paul R. Muessig, David H. Hall; Dennis R. Laack; Martha L. Hoppus; Barry O'Neal, Joint Technical Coordinating Group on Aircraft Survivability (JTCG/AS), JTCG/AS 97-M-008, October 1997

Pilot Verification Study Report, ENTEK/ABQ-94-0106-TR, February 16, 1994

System Safety Program Requirements, US MIL-STD-882C Notice 1 dated 19 Jan 1996, US Department of Defense

Authors' Background

Michelle L. Kilikauskas (principal author) is the Director of the Joint Accreditation Support Activity (JASA) at the Naval Air Warfare Center, Weapons Division (NAWCWD) at China Lake, California in the USA. She has served as the accreditation support agent for several international missile and military aircraft acquisition programs and was a participant in the International Test and Evaluation Steering Committee's Working Group of Experts on Verification and Validation. Ms. Kilikauskas is a frequent writer, speaker and teacher on the practical implementation of VV&A policy, principles and methodology.

Dirk Brade received a Diploma Degree in Computer Science from the Technical University of Darmstadt. Since then he has been employed by the Institut für Technik Intelligenter Systeme at the University of the German Federal Armed Forces where he has conducted basic research in VV&A for the Department of Studies and Exercises and the German Federal Office of Defense Technology and Procurement. He was also a member of the German delegation to the International Test Operating Procedures (ITOP) VV&A Working Group. Mr. Brade has given numerous papers on VV&A in various international forums.

Robert M. Gravitz has over 20 years experience in research, development, test, and evaluation (RDT&E) of major DOD weapons systems. His focus has been the application of M&S to material acquisition, program management, and systems engineering and systems verification. Mr. Gravitz has focused on planning of M&S verification, validation and accreditation (VV&A) activities in support of formal accreditation by senior decision makers. As Director of the Systems Engineering and Evaluation Technology Group within Aegis Technologies, Mr. Gravitz has current, hands-on experience with planning, costing and executing VV&A of complex, hardware-in-the-loop simulation systems. In this role he plans and directs M&S VV&A tasks for several Major Defense Acquisition Programs (MDAPs) for several government agencies. Resource planning and M&S VV&A execution programs presently supported by Mr. Gravitz include a wide variety of weapon system, hardware in the loop, testing and M&S systems. Mr. Gravitz has an MS in Systems Management from the Florida Institute of Technology and a BS from the University of Florida.

David H. Hall currently is the Manager of SURVICE Engineering Company's Ridgecrest Operation, under contract to the NAWCWD Survivability Division for analysis support services. Prior to his retirement from the Government in January 2002, he was the Chief Analyst of the NAWCWD Survivability Division, head of the Survivability Analysis Branches, and interim JASA Director. From 1992 through 1996 he was also the Joint Project Manager of the Susceptibility Model Assessment and Range Test (SMART) project, which developed and demonstrated Joint M&S VV&A and configuration management processes for DOD. At SURVICE he continues to participate in JASA VV&A support services to DOD systems acquisition programs. He is a frequent speaker, author and instructor on survivability and M&S VV&A.

Martha L. Hoppus is an accreditation support analyst with JASA responsible for coordinating the VV&A activities for the AIM-9X program. She also more recently has participated in M&S VV&A for the Rolling Airframe Missile program. Ms. Hoppus was instrumental in tracking and analyzing the cost of conducting V&V activities during the SMART program, using earned value principles and a detailed Work Breakdown Structure.

Ronald L. Ketcham is the Branch Head for the System Survivability Integration Branch at NAWCWD, China Lake, CA. He recently served as Chairman of the M&S Credibility Workshop held in Reno, NV (March 2002). Mr. Ketcham is also an Accreditation Support Analyst with JASA, and in that role, he is the Accreditation Support Agent for the Joint Strike Fighter (JSF) Program. Mr. Ketcham has extensive experience in software engineering and VV&A, and he gained much experience in M&S VV&A as Project Manager of the Analyst's Work-Bench (AWB) in the early 1990's.

Robert O. Lewis is a senior software engineer at the Boeing Company. He is currently part of the core Software Engineering Process Group (SEPG) responsible for developing and institutionalizing Software CMM and CMMI processes at the Huntsville site. He has developed three cost tools over the past seven years and is part of the Defense Modeling and Simulation Office (DMSO) team responsible for producing the new VV&A Recommended Practices Guide (RPG). He is also an ISO 9000:2000 Lead Auditor and a CMM-Certified Assessor and has participated in and led a number of Assessments. He has been very active in the Computer Simulation Conferences and Simulation Interoperability Workshops in the past and has authored more than 25 papers on VV&A. He contributed to the current RPG, several IEEE standards, and several other VV&A publications including DA PAM 5-11 and DODI 5000.61, and the VVT&E process for DARPA's Computer-Aided Education and Training Initiative (CAETI). He has taught more than 65 technical courses and seminars over the past 18 years and wrote a popular textbook on IV&V published by John Wiley & Sons.

Michael L. Metz is Director of Modeling and Simulation programs at Innovative Management Concepts, Inc., in Sterling Virginia. His current V&V responsibilities include: Technical Director for the Joint Warfare System (JWARS) V&V Program; Program Manager for the Military Traffic Management Command Transportation Engineering Agency's PORTSIM V&V Program; senior advisor to the Air Force Studies and Analysis Agency's STORM V&V Program; member of the US Navy Research Lab's WARCON VV&A Integrated Product Team (IPT); and a member of the DMSO VV&A Technical Working Group (TWG). He has been directly involved in the development, use, testing, and V&V of DoD simulations for 17 years as an Air Force officer and contractor. In his work on the DMSO VV&A TWG he has been a primary author of the DMSO VV&A RPG Millennium Edition and has developed new concepts for object-oriented simulation design verification and for the use of Subject Matter Experts (SME) in an objective results validation process.

Scope and structure of the paper

This paper was written by multiple contributing authors in order to include a wide variety of approaches for and experience with estimating the resources required for V&V activities. We begin by examining the factors that influence the required scope of V&V, including the complexity of the M&S being examined and the availability of prior information about the model and data to support validation. Since V&V efforts are (or at least should be) driven by the intended application of the M&S, the complexity and risk associated with the application also affect V&V resource requirements.

Several approaches will be discussed in an examination of the "state-of-the-art" of estimating V&V resource requirements, followed by a number of case histories from programs of various types and sizes within the Department of Defense. These include missile development programs, aircraft programs, and a major simulation system. Finally, an analysis of the state of the art and the program case histories will be presented, along with conclusions from the analysis and recommendations for further improvements and research.

Factors influencing the scope of V&V

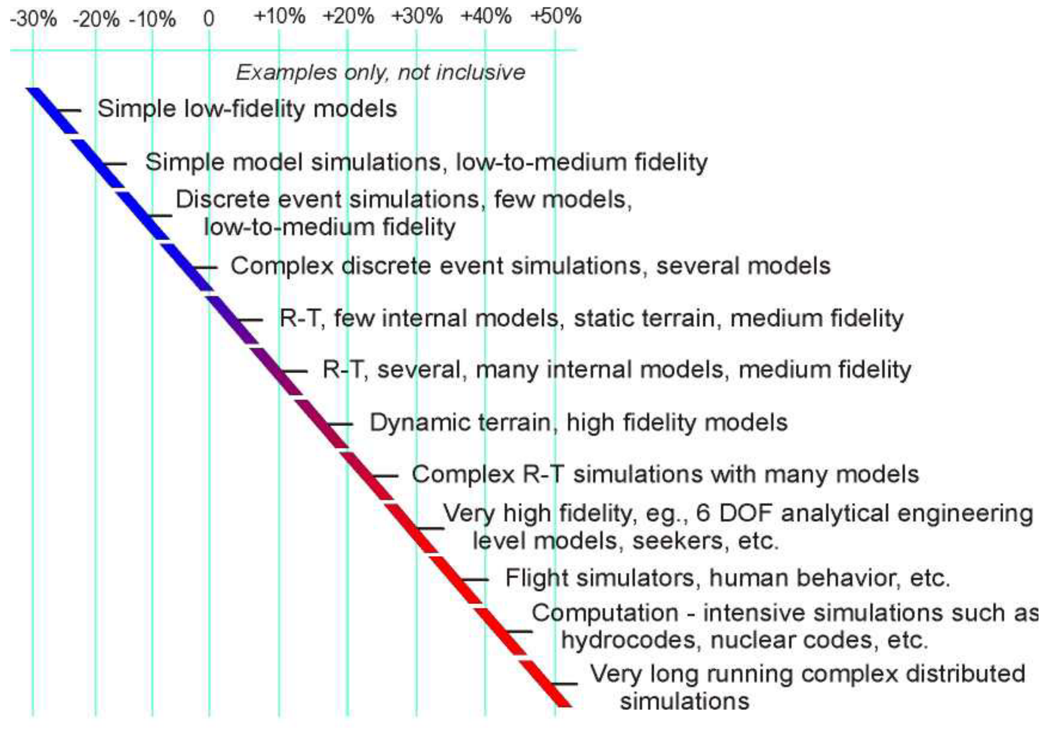
A number of factors influence the required scope of V&V activities. Some of these factors are associated with the simulation software (complexity, size, etc.) and some with the application of the simulation (the risk associated with using incorrect M&S results, for example). These factors complicate the process of estimating V&V resource requirements.

M&S Complexity

The resources required to support V&V for M&S generally are assumed to be a function of the size and complexity of the software. Most M&S development cost models use lines of code (LOC) or source lines of code (SLOC) as the basic metric for predicting software development cost. While LOC (or SLOC) does not necessarily correlate directly with model complexity, size usually is at least a general indicator of the complexity of the code and is assumed to also correlate with V&V resources required.

It is widely recognized that the complexity of M&S can vary by a factor of 20, 30, or more. In fact, there is no limit on the high end, except computer power and funding. To support this discussion, suppose a simple non-real-time simulation of System X might have 5,000 SLOC. On the other hand, a very highly realistic, real-time simulation of the same system with many attributes could easily have 150,000 SLOC. We commonly refer to this as complexity of the M&S. Robert Lewis, the designer of the Cost Estimating Tool (CET) discussed later in this paper, observes an interesting relationship between complexity of the M&S and complexity of the VV&A effort. Perhaps VV&A complexity would more appropriately be named “difficulty”, but complexity gets the idea across pretty well.

The CET uses the term complexity as a multiplier that increases or decreases the amount of V&V resources in proportion to the complexity of the M&S, but here is the catch: the VV&A complexity correction factor in the cost tool is scaled far differently than the complexity factor of the M&S. The reason for this is that the complexity (difficulty) factor for VV&A is largely mitigated because of the size of the software product that results. That is, the more complex the solution found in the M&S, the larger the software. Because the VV&A estimate is based largely on software size, it moves up or down to match, and the only other correction factor needed in the tool is one that classifies the M&S by gross characteristics as shown in the following figure. The range of correction factor in the figure varies between -30% and +50% based on complexity, with some examples shown of different types of M&S to give an idea of how the correction factor might be scaled.



Complexity Correction Factor in the CET

At the beginning of new M&S efforts when there is likely to be a high amount of uncertainty associated with the proposed SLOC estimate, it is recommended that the planners and estimators use an alternative way of deriving complexity. Although there are several ways to accomplish this, Function Points are a viable option since there is a known relationship between SLOC and Function Points for each given 3rd generation software language. Hypothetically, if one were to use the conversion of 57 SLOC per Function Point for Language "X", then on average, if there were 1250 Function Points one could expect the SLOC count would be around 71,000. If the SLCO estimate appears significantly higher, it could imply that the bidder's complexity estimate is higher as well (by the ratio of the difference). If this ratio were 20% higher, the user of the CET would enter the number 1.2 as the complexity correction factor. If it were lower, adjust downward accordingly. This is another way to estimate complexity and its affect on the VV&A effort. This concept is discussed in more detail in the Cost Estimating Handbook provided along with the CET software. The VV&A CET is covered more fully later in this paper.

Availability of information about the M&S

The resources required by a V&V effort are greatly affected by the availability of information about the model. For example, due to classification restrictions or just poor documentation, the available information about the simulation model itself may be limited. If for any reason the necessary information can not be reconstructed, verification activities will be much more difficult, and will likely be scaled back because of this lack of information.

Three cost factors must be distinguished here:

- Expense of buying information about the model
- Cost of reconstructing unavailable information
- Cost difference incurred when forced to replace a relatively "cheap" V&V technique with a more expensive V&V technique

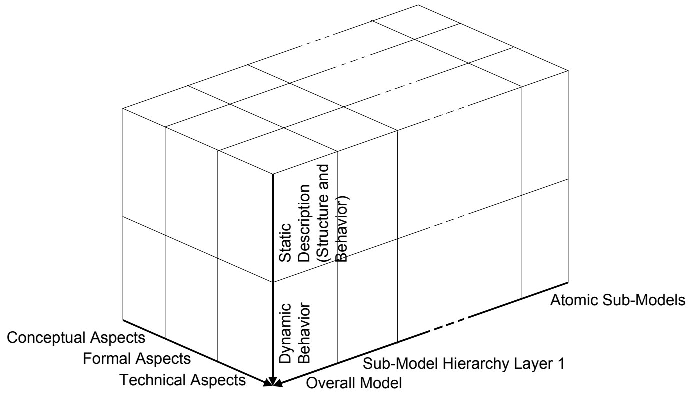
Different types of information address different aspects of a model. The type of information available influences the selection of V&V techniques and V&V sub-tasks. In the following discussion we introduce one possible framework for classifying information about a model, where model information is decomposed in three dimensions:

(1) Information about conceptual, formal-mathematical, and technical aspects of the model
(2) Information about the sub-models or components of the overall model
(3) Information about the model's static description and dynamic behavior

Information about conceptual aspects of the model provides insight into the ideas behind the model, e.g., the motivation behind the chosen system boundaries, idealization, and abstraction. It allows "cheap" validation, for example, by Subject Matter Expert (SME) opinion. Information about formal-mathematical aspects of the model allows us to understand how the structural and functional assumptions about the real system have been translated in computable equations for the purpose of digital simulation. It allows "cheap" V&V of physical equations or line-of-sight algorithms, for example. Information about the technical aspects of the model provides insight into the implemented Hardware/Software (HW/SW) solution, including the chosen programming paradigm, variable types, memory allocation, etc. It supports "cheap" examination of code structure and data management.

Similar to the system-of-systems decomposition of a real system, in many cases the model can be considered as a model-of-models. The more that is known about the sub-models, their hierarchy, their interactions and interdependencies, the more knowledge can be gained concerning the overall model. In the following, the components which are not decomposed further are called "atomic sub-models". Sub-models created by composition of atomic sub-models are allocated on sub-model hierarchy layers, with the overall model as the highest level of integration. The hierarchical (de-) composition allows cheap bottom-up and top-down examination.

A static description of the model allows us to gain knowledge about the structure and behavioral rules of the model (the top box in the figure below). Analyzing the model can be "cheap" if we have a static description available. Observation of changes of model output as model inputs and internal states change gives insight into its dynamic behavior (bottom box). This type of sensitivity analysis is "cheap" and provides a great deal of information about the correctness of the model's behavior.



A 3d-framework for classification of model information

The three dimensions introduced above create a 3d-framework for the classification of information about a model, as depicted in the figure. With this characterization it is possible to differentiate information about the model by information subspace (the boxes in the figure) per the discussion below. By having V&V activities associated with each information sub-space, and by understanding the transitions between the sub-spaces, we can then identify which V&V activities are cost-effective within that framework.

V&V in the Information Subspaces

Examples of V&V activities that might be conducted in the information subspaces pictured above include:

- Input and output data recorded during a run of the executable model (such as sensitivity analyses) belongs to the information subspace "dynamic behavior of the overall executable model". Statistical analysis of this type is relatively inexpensive. However, at this level (within this subspace) no conclusion about the structural appropriateness and correctness of the model can be drawn no matter how much is applied in the way of resources.
- Recording the I/O behavior of federates when executing a federation of simulations belongs to the information subspace "dynamic behavior of the executable models on sub-model hierarchy layer 1"…
- If an analyst thinks through the formal-mathematical description of the behavior of an entity within the model, this type of V&V activity belongs to the subspace "dynamic behavior of the atomic sub-models".

Dependencies between the Information Subspaces

Of course, these subspaces are not, in general, independent of one another. Some examples of dependencies between these subspaces include:

- If one has a compiled binary executable version of the M&S on a computer platform, information about the dynamic behavior of the overall model can be easily created and the technical, formal-mathematical, and conceptual aspects can be more easily examined.
- However, in that case the static description of the overall model is not necessarily easily available, as one would have to disassemble the binary code to learn about the static description and the technical aspects of the overall model, and then reverse-engineer the code to understand the formal-mathematical aspects of the overall model description and the underlying concepts.
- Without any way to delve into the code (difficult in binary code), no information about the dynamic behavior of the sub-models would be available, for example. If the analyst had access to the UML description of the model, analysis of the formal-mathematical and technical aspects of the overall model would be possible, even for the below sub-model layers.

The availability of information about the model heavily influences the possible choice of cost-effective V&V activities. We are interested in validating the behavior of the overall executable model, but since this more often than not is not directly possible, we must find a path through the information subspaces and assess the resource requirements for the V&V activities that are achievable in each of the subspaces.

Availability of Validation Data

A counterpart to dynamic behavior data about the simulation is the dynamic behavior of (a representative of) the real system that is being simulated. These data are necessary to conduct simulation validation activities. The availability of these test data or other data with which to compare M&S predictions is generally the biggest driver of M&S validation cost. However, the costs of obtaining those test data are not very often associated with validation in terms of cost tracking. This is because most test events that provide data for M&S results comparisons are usually done for other reasons, such as operational test events for developmental systems (the tests are done to demonstrate the system under test, not the M&S). Open air range test costs are in the millions of dollars per test, and thus the cost of testing makes M&S validation very expensive if not planned for very carefully. The result is that it's hard to predict what it will cost to do validation, especially if you can't find someone else to pay for the testing.

And if you do find someone else to pay for the testing, another complication for costing out validation is the release-ability of data. Often the program paying for the test is very sensitive about letting others have or use their data for other purposes. It doesn't help for someone to have the data you need if can can't get access to it, or if you can't release your validation results to others. Classification issues can cause similar problems for M&S validation efforts. Either of these issues will raise the cost of obtaining validation data, either through efforts required to "sanitize" the data of any elements that are objectionable to the program, or to process the data in such a way that programs cannot be identified, etc.

Also, because many validation efforts use data collected for other purposes, often the data collected are not usable for M&S validation. This is because specific pieces of data necessary for model validation often are not collected during the test, not collected with sufficient precision, the instruments were not calibrated, or the instrument pack fell off the test article, etc. Thus careful and detailed planning is required prior to testing to ensure that M&S validation data requirements are included in the test plans.

Application Complexity

Application complexity has to do with the way that M&S results are used by the "customer" of the M&S. If the M&S are part of a large, highly integrated process using analysis, testing, training, or other "live" participation in the overall process, then it may be difficult to separate out requirements for V&V activities for the M&S. If they are used as just a piece of the overall customer's application, then determining fidelity or accuracy requirements may be very subjective, since the M&S results cannot be easily separated out of the overall process. It will be difficult to say, for example, that this M&S result must be "accurate to within X%", if the M&S result cannot be easily extracted from the overall application. This may be true even if live participants are not involved; determining accuracy requirements for simulation federates can be difficult, for example, for various reasons. And complex interactions between linked M&S can cause complex interdependencies among M&S outputs.

Application Risk

High risk applications (where a wrong answer can produce catastrophic results, or damaging results with high probability) require more resource expenditures on V&V than do low risk applications. Both the JASA and U.K. approaches described later in this paper address the effects of application risk on V&V tasking requirements (and the associated resource requirements). The approach developed within the ITOP group was also risk-based. Application risk is probably the greatest driver of V&V resource requirements.

Whenever simulation results influence the real world, there is a risk incident to their use. Simulation results must only then become the basis for a decision if they are sufficiently credible with respect to the impact the decision will have and the influence of the simulation results in the decision making process. For example, wrong behavior learned in a training simulator may lead to severe damage or loss of the real system during its operation by the wrongly trained person, or it may even lead to their death. The following discussion focuses on the consequences for simulation based decision making.

Typically, the risk R_{Sim} incident to a simulation based decision is defined as the product of the probability $P_{Sim}(E)$ of making the wrong simulation based decision and the worst case impact $I(E)$ of the wrong decision. This can be expressed as

$$R_{Sim} = P_{Sim}(E) \cdot I(E) \qquad (1)$$

The impact $I(E)$ of the wrong decision may be directly derived from the area the decision affects, and is completely independent of the information the decision was based on. If simulation results are used, for example, to reorganize the traffic in a city, the worst case impact $I(E)$ may be a traffic breakdown during rush hour, with all its consequences for health and the economy. If the decision may result in death or massive financial loss, the impact of the decision is considered to be high. Otherwise, if the injury or financial loss is negligible, the impact is low. The judgment as to the severity of the worst case impact $I(E)$ of a wrong decision is not up to the people involved in developing the M&S and conducting V&V, but to the user of the M&S. Only the user understands his or her application well enough to determine the impact of a wrong decision. As an externally provided constraint, $I(E)$ can be thought of as "constant" with respect to V&V, since the quality of the simulation results does not influence the worst case impact at all.

The probability of a wrong decision $P_{Sim}(E)$ depends on the appropriateness and correctness of the information base on which the decision is made. This information base may be completely or partially gained through the use of M&S results. From the perspective of those involved in M&S, the probability $P_{Sim}(E)$ of a wrong simulation based decision can be decomposed into (1) the probability $P(E|O_E)$ that wrong simulation results lead to a wrong decision, and (2) the probability $P(O_E)$ that the simulation results actually are wrong. This can be formulated as

$$P_{Sim}(E) = P(E \mid O_E) \cdot P(O_E) \qquad (2)$$

Depending on how much simulation results influence the decision, the appropriateness and correctness of the simulation results influence $P_{Sim}(E)$ more or less. In other words, if the simulation results are not considered much during the decision making process, their impact on the decision is respectively low, which can be expressed by a low $P(E|O_E)$. If the decision is completely simulation-based, the impact of the simulation results is high, resulting in a high $P(E|O_E)$. If other sources of knowledge are consulted in addition to M&S results, the impact of the M&S results depends on their weight during the decision making process.

Treating the share of simulation gained knowledge on the decision base as an a priori requirement, $P(E|O_E)$ is not influenced by V&V and therefore can be considered a "constant." The influence of V&V activities is to reduce the probability of negative influences of M&S results on the overall decision making process by reducing the probability $P(O_E)$.

The decision maker has to accept a residual risk R_{Max} when making a decision. The <u>acceptable</u> residual risk R_{Max} is treated as constant with respect to V&V, as V&V does not modify it. However, V&V activities can drive the actual residual risk below that acceptable value, R_{max}. If $R_{Max} \geq R_{Sim}$, i.e., the risk R_{Sim} of making the decision on the available information base is smaller then or equal to the maximum acceptable residual risk R_{Max}, the model or simulation results may be used with acceptable risk. Using equations (1) and (2) this can be formulated as

$$R_{Sim} = P(E \mid O_E) \cdot P(O_E) \cdot I(E) \leq R_{Max} \,. \qquad (3)$$

With all terms being constant from the V&V perspective except for $P(O_E)$, the maximum acceptable $P(O_E)$ can be expressed as

$$P(O_E) \leq \frac{R_{Max}}{P(E \mid O_E) \cdot I(E)}. \quad (4)$$

Unfortunately this discussion is still academic. Although the equations create the impression that it is possible to calculate quantitatively the functional dependencies between the risk of simulation use and required V&V, this impression is misleading. Due to the lack of clearly identified and commonly accepted measures of model quality, it is not reasonably possible to calculate $P(E|O_E)$ or $P(O_E)$. But the following (intuitive) qualitative conclusions can be drawn from the above discussion:

- The lower the maximum acceptable residual risk, the lower the acceptable $P(O_E)$.
- With increasing $P(E|O_E)$, $P(O_E)$ must decrease.
- If $P(O_E)$ can not be reduced under a certain threshold, $P(E|O_E)$ or $I(E)$ must be decreased and/or R_{Max} must be increased.

As a consequence of this uncertainty in estimates of impact and risk, we will see that the risk-based approaches that we discuss in this paper make use of subjective judgments, as opposed to actual calculations of risk. All of the approaches that are used in practice categorize risk probabilities into low, moderate and high, and impacts into catastrophic, severe, moderate, negligible, etc.

Software Risk and Uncertainty (R&U)

A detailed example of such an approach is embedded in the Cost Estimating Tool (CET), described later in this paper. Users of the cost estimating process defined in the CET are required to answer a set of 15 questions intended to establish Risk and Uncertainty (R&U) factors for the M&S and VV&A program being estimated. In most cases, reuse of a legacy product that has some VV&A history and few accumulated problem reports presents only small to moderate risk, but as the amount of modification increases, so will the risks and uncertainty. Thus, the user of the CET is required to select a value that he or she feels correctly represents the risk or uncertainty associated with each question. If undecided or normal risk is assumed, the value for that question stays at the mid-point of 50 points. Mid-point values have already been established and pre-set for each use case and type of M&S. If the analyst believes from the details of the question that it represents greater than normal risk, a value above 50 is selected using a slider. If it is determined to be lower, a value of less than 50 is selected. Most R&U values fall in the range of 35 to 65, and to stay within these numbers is prudent unless there is a compelling reason to go beyond them. This is especially true in cases where the VV&A history and user experience are both good. Users should have very strong reasons for selecting values above 70 on the high side or below 30 on the low side.

A "Risk and Uncertainty Profile" is illustrated in the following table. This table provides notional levels of V&V required as a function of those 15 factors evaluated by the CET, factors which identify the risk and uncertainty associated with using results from the M&S being evaluated. Normally, if an R&U factor "fails" one criterion in the Risk and Uncertainty Profile (that is, requires "High" V&V), that question can be bumped up about 5 points; if it fails two, perhaps 8 to 10 points; and so on up the line. Likewise, if it more than meets the criterion, reduce it by 5 points per criterion, but be careful not to overdo the reductions or increases since they can have a large effect on the adjustment of the final cost estimate for the V&V effort. In as-is legacy applications, it is not unusual to have most of the 15 questions remain at the mid-point. Only adjust those that you feel are really outside the normal range, in either direction, since the cost model has already been normalized to an average risk value and could be overly compensated by an overly zealous approach to risk and uncertainty. For example, it may be that an M&S simply does not have an adequate conceptual model. In this case, Question 15 could be increased to 60 or even 65, depending on just how poor the conceptual model actually is. In addition, Questions 3, 4, and 7 (and possibly others) are candidates for small increases (probably 5 points or more) since there can be minor problems with the documentation, requirements, input data, and perhaps additional training required to operate the M&S.

As the 15 questions are completed, the R&U factors are entered into a matrix, which is part of the automated tool, where they are matched to VV&A activities. Each VV&A activity may be affected by multiple R&U factors. These intersections inside the tool's matrix are called "hits." This allows the highly critical factors to have a proportionally greater number of hits and, thus, have a greater affect on the estimated VV&A level of effort. This tunes and weights the cost model to the risks and uncertainties of each particular M&S program. The values in the matrix are cross-multiplied and produce counts that are adjusted by R&U.

Factor	Low V&V	High V&V	Definition	Attribute Considerations
1. Development Program	Well Defined	Poorly or Undefined	Measure of the completeness, reasonableness, and adequacy of the programmatics	• Adequacy of funding profile and plans • Schedule reasonableness and attention to constraints • Resource availability when required • Qualified mature developer (ISO, CMM, etc.) • Adequacy and availability of support infrastructure
2. Technical Risk	Low	High	Measure of the technical risks (SW, HW, network, etc.) associated with key implementation factors	• Reasonableness of the technical approach • Implementation feasibility • Dependency on new technology • Understanding of the physical properties • Maturity of the domain (subject of the program)
3. M&S Documentation	Adequate Complete	Inadequate, Incomplete	Measure of M&S documentation adequacy and completeness	• Formal traceable requirements • Well developed & documented design • Well commented & maintained code • Comprehensive test plans & procedures • Adequate User & Operator Manuals
4. System Requirements	Stable	Unstable, Incomplete	Measure of the stability of the system requirements with respect to development	• Well chosen development paradigm • Frequency of change in requirements (stability) • Documented requirements (representations, models, specs, etc.) • Verifiability, modifiability, testability, and traceability • Completeness, consistency, and correctness of the requirements • Functionality, performance, design constraints, etc.
5. Maturity of Technologies	Mature	New, Unproven	Measure of the maturity of the technologies (SW, HW, domains) to be used or applied in the program	• Identified technologies • Proven M&S concepts • Support for advanced technology • State of current technology • Insertion approach, frame work, or implementation method
6. Application Reusability	High Percent Available	Low Percent Available	Measure of the availability of existing applications (GOTS, COTS, In-house) that meet the needs within the program boundaries	• Degree of reusability of SW & HW • Similar applications available • Legal Constraints (software rights, royalties, license fees, etc.) • Integration of existing applications • Reliability and integrity of the reusable application • Available within schedule and budget constraints
7. Data Usability	High Percent Available	Little or None Available	Measure of availability of existing data sources that meet program needs	• Similar applications exist that use the same data • Existing data & authoritative sources are available • Data constructs, schema, design, etc. are available • Available within schedule and budget constraints

Risk and Uncertainty Profile

Factor	Low V&V	High V&V	Definition	Attribute Considerations
8. Communications	Proven	Unproven, Uncertain	Measure of the uncertainty of the communications needed for the program	• Network complexity • External and internal network dependencies • Performance required by application (latency, bandwidth, etc.) • Interfaces between devices • Perceived understanding of communications needs
9. Fidelity	Low	High	Measure of the accuracy of the perceived functions within the program	• Features needed to represent real world adequately • Performance and functionality required • Details needed at various levels • Adequacy of graphical user interface & visualization • Reliability and accuracy of algorithms
10. Personnel Requirements	No Training Required	Detailed Training and Skill Level Required	Degree of training, skills, proficiency required to use or operate the system	• Any training required for V&V staff • Form of training required - classroom, on-the-job, etc. • Formal training, certification, or licensing for users • Lead times, conflicts in training and using the system, etc.
11. User Environment	Simple	Complex	Measure of the complexity of the user's environment, (outputs, representations, actions)	• Measure of stability • Complexity of operations, screens, output data formats, etc. • Special skills to use simulation • Special skills to interpret output
12. Architecture	Simple, Monolithic	Complex, Distributed	Measure of complexity of architecture, both physical and functional	• Size and complexity of the configuration • Number of distributed processes and functional elements • Physical separation among parts • Degree of coupling among parts • Amount of real time processing required
13. Domain Complexity	Simple, Non Real-Time, Non-Critical	Complex, Embedded, Real-Time Critical	Measure of domain complexity and criticality	• Complexity of behaviors and control mechanisms • Time-criticality of simulation (real- vs. non real-time) • Required accuracy of system • Complexity of algorithms
14. Verification and Validation, Testing Support	Cooperative, Timely	Uncooperative, Slow, Non-responsive	Measure of developer's willingness to cooperate, share data, resources, etc.	• Maturity of development and test environments • Timeliness of response to requests for shared testing • Continuous flow and exchange of artifacts in both directions • Cooperative configuration management
15. Formal Conceptual Model	Available & Adequate	Non-existent or Inadequate	Measure of adequacy and correctness of Conceptual Model	• Represents M&S functionality, behaviors & performance • Includes key algorithms • Links to requirements & authoritative sources • Usable and understandable

Risk and Uncertainty Profile

Accreditation Authority Requirements

The amount and type of evidence required by the accreditation authority to come to an accreditation decision also drives V&V resource requirements, especially since the accreditation authority is usually the one paying for the V&V effort in the first place. The amount of evidence required by a "sponsor" may or may not seem logical to M&S practitioners, since it usually is driven partly by policy, partly by application risk, partly by previous experience, and often by preconceived opinions about M&S (for or against) on the part of the program manager. These factors, while subjective at best and non-measurable at worst, must be considered in any prediction of V&V resource requirements. It is an unfortunate fact that VV&A requirements, and thus resource requirements, are more often driven by political, rather than technical issues.

M&S Task Accounting

Projecting the resources required for V&V based on historical data is made difficult by the accounting procedures that are followed by most programs for tracking M&S costs in general. How V&V tasking is accounted for by a program can make it difficult to figure out the cost of VV&A as part of an overall M&S development program or of a larger program for which simulation development and use are part of the larger goal (like building and fielding a weapon system). What activities do you count as "simulation/software development" and what do you count as V&V? If a software engineer spends an hour coding and an hour desk checking and running test cases, is it an hour of development time and an hour of V&V, or are desk checking and running test cases part of development?

Going back to the discussion on software V&V vs. simulation V&V, is verification part of development but validation a distinctly "V&V" cost? If I hold a peer review within the development team, is that part of development? If I invite outside experts who are not already on the development team's payroll to participate, do I now consider the cost of the experts as a V&V cost, not a development cost? How about model documentation? Is that part of development or is that a separate V&V cost?

In terms of simulation validation, the accounting problem can be even more extreme. If doing preflight predictions is a normal part of preparation for a missile flight test and the predictions are used as one consideration in a mission readiness review, is the time required to do the preflight predictions a V&V cost or a test cost? If it is standard practice after a test firing to run the simulation with the actual flight test conditions and carefully examine the telemetry (TM) and the simulation predictions to look for anomalies and to understand what happened in the flight, and it turns out that the post-flight analysis is useful for validation, where do you log the cost of doing the post-flight analysis? Is this a V&V cost, or is this part of the test program? On one program, for example, the accreditation agent added some tasking to the overall M&S plan in order to do more formal documentation of the post-flight analyses and examine a few more parameters than originally planned. In that case, the extra documentation work was considered to be V&V tasking, but the actual post-flight analyses and comparisons between test data and model results were counted as part of the T&E cost. So what were the M&S validation resource requirements in that case?

If a Hardware-in-the-loop (HWIL) facility is used to examine the behavior of a missile under test before it is fired and some good validation of the HWIL comes out of that effort, where do you log that cost? If you use data from an exercise to do some validation on a mission level simulation, do you consider all the costs of the exercise as part of V&V? If so, what costs do you consider? Do you count the labor of the military guys who drive the tanks even though their salary isn't directly charged to the exercise?

For simulations of real objects that are under development, it is exceedingly difficult to figure out which costs go into the "item" development bin, the "M&S" development bin, the "V&V" bin, etc. The more that good software/simulation development practices are followed, the harder it is to sort out development costs from V&V costs (because software V&V is integrated into the software development process). The more the simulation is used as a key tool in the development of the object, the harder it is to sort out M&S validation cost from the cost of developing and testing the object. These observations come out of direct experience with aircraft and missile development efforts, but engineering level models that support developing a car or a tank or some other item are likely to be much the same.

Practitioner Expertise

During the Susceptibility Model Assessment with Range Test (SMART) program, five engagement level simulations were used to demonstrate a VV&A process that the SMART project developed. These models varied between 30,000 and 100,000 lines of code, but the V&V tasking and resources expended on each was about the same. This may have been partly due to the fact that the M&S examined by SMART were all "legacy" codes which had no documented conceptual models; part of the V&V tasking involved developing those conceptual models by "reverse engineering" the code, using expertise still available at the developer's facility and other resources. And none of the five models were entirely verified or validated, due to cost and time limitations. Nonetheless, one overall conclusion from SMART was that the experience and expertise of the people doing the V&V was much more of a factor in determining resource requirements than the size of the code. This may not carry over into very large or complex codes (much larger than 100,000 SLOC, for example), but it still is likely to be true that the subject matter expertise of the people involved in doing V&V will be a driver of V&V cost, especially for very complex codes.

Estimating V&V Resource Requirements: The State of the Art

Existing policies address VV&A in many communities that use modeling and simulation, with the U.S. Department of Defense being a prime example. These policies generally acknowledge the need for tailoring the processes outlined in policy to fit the needs of specific situations. In particular, VV&A processes often are tailored based on the risks associated with using incorrect M&S results: the higher the risk, the higher the confidence the user must have in the M&S results. The approaches described below provide guidelines that could be helpful in that risk-based tailoring process. These risk-based approaches help to plan for which V&V and related tasks need to be done, which in turn drive the cost (and schedule) of V&V.

Risk Based Approaches - Managed Investment

Within the Department of Defense (DOD), accreditation is broadly defined as the official certification that a model or simulation is acceptable for use for a specific purpose[10]. Individual M&S and agency accreditation plans will be unique, but the M&S V&V activities selected for execution should be intended to provide essential, fundamental information about the simulation to support M&S accreditation decisions.

As a consequence, accreditation should be the primary objective in the definition of the M&S V&V activities. The goal is to establish that the M&S produces realistic, unbiased, credible measurements of performance parameters when operated within a specific domain of scenario and environmental conditions for it to be acceptable (accredited) for use. The essence of M&S V&V therefore is establishing the degree to which decision-makers may have confidence in the results of studies, analyses, tests, and training exercises conducted using the M&S; e.g. that the M&S sufficiently reflects the real world[11] from the perspective of the intended use. The scope of evidence applicable to that determination includes:

- M&S development activities, test and integration activities and the supporting M&S developmental, test and integration documentation
- Configuration management (CM) processes for the M&S and the CM supporting documentation, including trouble reports, deviations, waivers; notices of revision
- VV&A activities that have been completed to date for the M&S and formal documentation of the results

Thus, much of the VV&A planning and execution process consists of generating, organizing, and reporting the evidence developed or originated in M&S development, test, and configuration management activities, as well as M&S VV&A activities documented in an auditable form for

[10] Department of Defense Directive (DODD) 5000.59, "Department of Defense Modeling and Simulation Management"

[11] DoDD 5000.59 defines Real World as: "…the use of actual components in authentic experiments and exercises that include live weapons, real communications, actual sensors, actual and mock engagements, etc. It includes the real environment, observed phenomena, and the randomness of natural occurrences. It can and often does include humans in the decision-making process, which adds the unpredictable nature of the cognitive strategy and outcome".

the Accreditation Authority's use. Each of these inter-related activities can contribute significantly to establishing the foundation for user acceptance and formal accreditation.

But, "What V&V should be done?" and "How much will it cost?" are implicit to the M&S VV&A planning conundrum wherein financial resources for M&S VV&A execution almost always comes from, and competes with, all the other requirements levied on the financial resources of the M&S project.

A "Managed Investment" strategy for M&S VV&A provides for a deliberate and progressive outlay of resources to garner the information necessary to support M&S accreditation decisions. Thus, an actual V&V evaluation suite can be identified which is the most cost-effective within the space of possible candidate activities. This sub-domain then constitutes an optimal investment in V&V activities and products. This is essential for Department of Defense simulation programs in which nearly all VV&A projects operate within an economically constrained environment.

This strategy has been successfully demonstrated by AEgis Technologies in supporting the M&S programs of several Major Defense Acquisition Programs (MDAPs). The practical application of this approach will be illustrated in one of the case histories presented in this paper.

Managed Investment Defined

As a first step, let us define the term. Managed Investment is the execution, from all the possible candidate V&V activities, of a carefully selected subset of V&V activities:

1) ***Offering the "best return on investment"*** by providing the essential information necessary from V&V findings, and
2) ***Providing the required evidence*** supporting the Accreditation decisions of Service and DOD Accreditation Authorities.

In this approach, cost is considered as an independent variable during the selection and execution of VV&A assessment activities. An optimal subset of VV&A activities can then be chosen based upon the:

- Assessment data needs of the Accreditation Authority
- Realities of the program (schedule)
- Fixed resources (budget) available for assessment and V&V activities

Thus, for any situation or simulation, the Managed Investment Strategy asks the same question, "How do I know if I'm doing, or have done, enough of the right V&V?" There is no objective or rigorous calculus for answering this question, but in attempting to answer it, we will identify what to consider and we will build a heuristic framework around the question.

Doing What Makes Sense

We want to distinguish between this "Managed Investment" strategy and the more commonly marketed "Exhaustive M&S VV&A Strategy" that says we should do every kind of verification and validation technique we can think of, fix every error we find, and that we should keep assessing the simulation until the sponsor pries it from "our cold, dead fingers", or "runs out of money". Even when the planned V&V activities themselves have not been completed due to resource constraints, the M&S program becomes "exhausted" in the process of discovering the strategy itself is not executable.

Some VV&A practitioners appear not to know when to say "when", or even care that an "Exhaustive M&S VV&A Strategy" is unworkable – they continue to advocate this strategy as a means of obtaining what they believe is their "fair share" of the M&S budget.

But, most M&S VV&A practitioners know exhaustive VV&A is irresponsible and unsustainable even if they are unable to articulate a viable alternative. Certainly, nearly every Program Manager, systems engineer, and software developer knows that an exhaustive software testing, M&S V&V, (and IV&V) effort is impossible, in both principle and practice, due to financial and schedule constraints. However, that doesn't stop some M&S programs, on the advice of VV&A practitioners or outside consultants, from pursuing an "Exhaustive M&S VV&A Strategy" which has the unintended consequence of ruining the program financially, and leaving the simulation sponsor with a deeply-held conviction that V&V is a "black hole" draining program resources better spent on M&S development (or avoiding M&S in the first place, and concentrating on other program requirements).

Although such an "Exhaustive M&S VV&A Strategy" usually fails, a "Managed Investment" strategy often can be implemented that is focused on the timely completion of selective M&S VV&A activities that, although reduced in scope or level of effort, are still sufficient to support an accreditation determination.

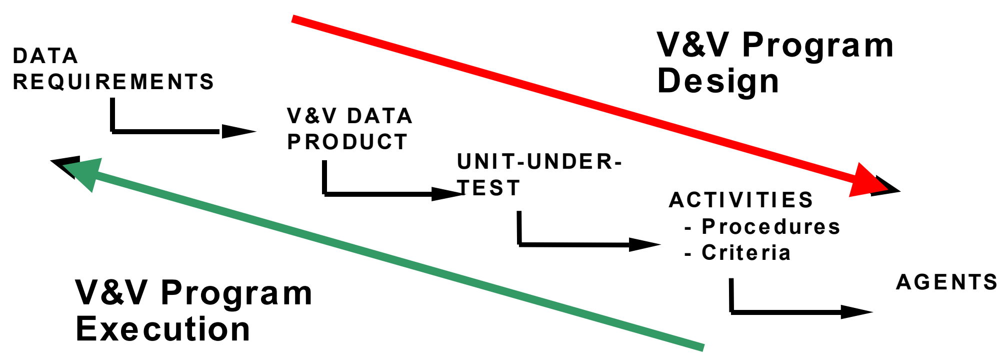
The challenge in moving to a "Managed Investment" strategy is being able to articulate (and perhaps quantify) to decision makers and budget analysts why certain M&S V&V techniques make sense for some simulations, but are unnecessary or wasteful for others. Let us show you how to make such an argument.

Requirements Driven VV&A Program

Most of us understand that *requirements* for any M&S VV&A program may be driven from the top-down, while V&V program execution can be built from the bottom-up. This chestnut of systems engineering is novel only insofar as its implementation is taken seriously.

But what are these "requirements"? Are these "simulation requirements", "software requirements", or "design requirements" which need to be verified or validated? The short answer is "No". You might ask, "Why is that?" Well, the goal of any V&V activity is to achieve the appropriate qualification of a given tool for a given purpose by a particular agency. It therefore makes sense to start by identifying the basis of such a judgmental decision, inferring the forms of evidence sufficient to support a positive outcome, and further deriving the means to generate that data. Then you can prepare for review and deliberation such evidence as is necessary and sufficient to support the accreditation decision of the potential Accreditation Authority. Thus, the focus is not compliance to system, software or design requirements for the simulation or the software; the focus is the accreditation information required to support the Accreditation Authority determination to accept the M&S for use.

This M&S accreditation information requirements-driven process is indicated in the illustration below, wherein accreditation information data requirements flow downward to V&V data needed for the accreditation Authority to make a prudent accreditation decision and avoid a Type II error[12]. Implementation is through V&V agents (possibly including Systems Engineering and Technical Assistance (SETA) contractors, V&V contractors, Operational Test Activities (OTA) and OGA) executing a suite of assessment activities for particular M&S, or units-under-test (UUT) to generate the necessary V&V data products and information to support user and Accreditation Authority acceptance decisions. Particular steps in the ladder-down of requirements are discussed in detail below.

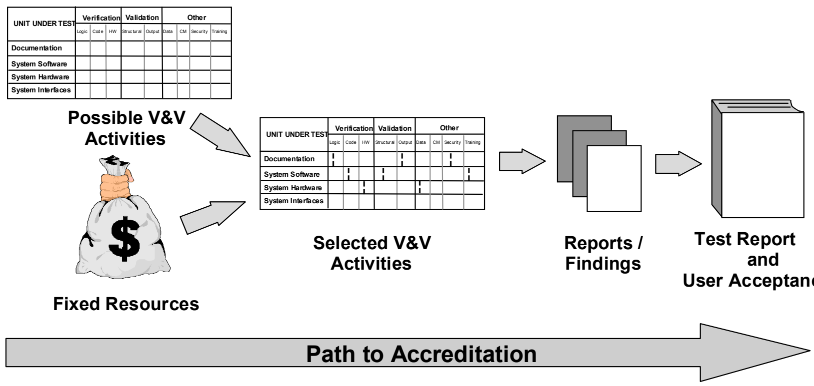


Accreditation Information Data Requirements Flow-Down

Difficulties exist, of course, in anticipating *all* user data requirements and information needs, associated acceptability criteria, and preferences for evidentiary support. Still, in those cases it is expedient to assume those positions and build a program of action while preserving an audit trail of requirements. This serves as a ready basis for the tailoring of a practical, effective, and reasonably low-risk strategy for any M&S VV&A program.

Managed Investment attempts to select the most cost-effective subset within the space of possible V&V activities, resulting in a near optimal V&V investment of the fixed resources available for the M&S. This is graphically illustrated below.

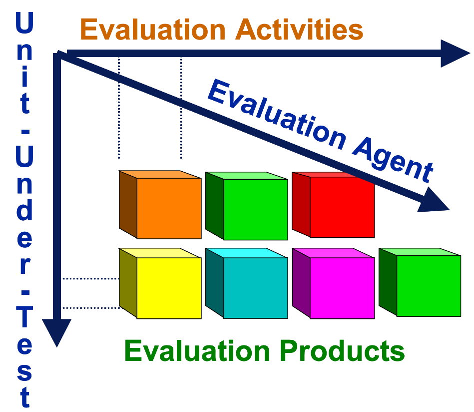
1. A Type II Error occurs when invalid simulation results are accepted, even though they are not sufficiently credible. Committing this type of error can be catastrophic. This is especially true if key decisions are based on the M&S results. The probability of committing this type of error is called Model User's Risk. The Type II Error is best avoided by completely understanding the application requirements and carefully considering the simulation results.



Managed Investment Provides the Resources for the Critical Assessment Activities Needed For An Accreditation Decision

VV&A Evaluation Activity Space

Another significant concept which is key to the "Managed Investment" strategy and supports M&S VV&A program definition is also a familiar one - it is the systems engineer's multi-dimensional view of the enterprise whose dimensions exhaust the important attributes of the conceptual space. Here we posit an "evaluation space" whose (relatively orthogonal) dimensions consist of: 1) V&V activities, 2) V&V agents, and 3) units-under-test. The points or cells in this evaluation space represent the V&V data products that are produced when a V&V agent carries out a V&V activity to evaluate a particular unit-under-test. This space is indicated (imperfectly) in the Figure (and it hearkens back somewhat to the Information Subspaces discussed under the section headed "Factors Influencing the Scope of V&V").



M&S VV&A Evaluation Activity Space

Each dimension is described in detail in the following paragraphs, after which we will illustrate the use of this construct in mapping-out and populating a practical VV&A plan-of-action.

The M&S VV&A data products (the cells in the figure) comprise the evidence for user acceptance and formal accreditation. The evaluation product requirements are identified through development of a select set of candidate activities that are coordinated with potential M&S users and the Accreditation Authority. The anticipated classes of data products that may be considered in the accreditation decision include:

1) ***M&S VV&A Administrative Documentation***; i.e., V&V Plan, Accreditation Plan, summary V&V Reports; Acceptability Assessment, and Accreditation Recommendation;

2) ***M&S Development Documentation***; i.e., System Specifications, System Design Documents, and Software Requirements Specifications and related documentation, CM Plan, User's Guide, Training Materials, etc.);

3) ***M&S Evaluation Documentation***; i.e., test and V&V assessments, including Integration and Test Plans, Software Test Reports, and V&V reports generated as a consequence of executing the VV&A Plan.

4) ***Other Technical Reports and Data*** generated by other evaluations (Requirements Analyses, CM Reviews, Subject Matter Expert Evaluations, V&V Analysis Reports, etc.).

Units-under-test (UUT) are those components of the M&S to which V&V evaluation activities are applied and upon which judgments are made. For most simulations, several entities (sub-models, objects, etc.) exist which needed to be verified and validated to establish user confidence and establish credibility of the M&S. Candidate UUT components, or facets of the M&S to be considered, are indicated by the enumerated items in the Figure below.

Naturally, the design of V&V exercise activities depended on the nature of the UUT (for example, we could validate analytical models, verify code, validate system models, certify (validate) input data, etc.). Because the variety of entities that comprise a particular M&S may be quite large, and because the items themselves may be so disparate, a variety of evaluation procedures may be required, so it is imperative that we explicitly identify each UUT and the associated V&V activities within the VV&A Plan.

SYSTEM SOFTWARE
- System Configuration Code
- Framework
- Common Model Set Code

SYSTEM CAPABILITY
- Experiment Preparation
- Experiment Execution
- Experiment Analysis

ANALYSIS TOOLS

SYSTEM MODELS
- Common Model Set Algorithms
- Specific System Representations (SSRs)

DATA
- Rulesets
- Characteristics Data
- Gameboard Data
- Scenarios

DOCUMENTATION

Candidate UUT

Evaluation Activities are selected V&V techniques and assessment procedures to be applied to relevant UUT to generate V&V data of interest and upon which acceptance criteria can be established. Several considerations are pertinent to M&S V&V and assessment activity planning in general.

First, defining evaluation activities requires us to carefully specify the evaluation procedures and criteria. Second, the details of activity specification effectively define the V&V program: activity flow and duration determine the program schedule, and the choice of assessment activities determines the level-of-effort (LOE) and associated resource requirements.

Finally, *every* V&V evaluation activity should be required to yield a valuable data product that facilitates user understanding and acceptance. Classes of potential evaluation (V&V) activities include those in the Figure.

Evaluation Agents are those principals that:

- Serve at the behest of the M&S Program Manager, Simulation Sponsor and, or other Accreditation Authorities;
- Execute the planned V&V and assessments activities; and
- Generate reports that serve to document the V&V activity.

VERIFICATION:	VALIDATION:
- Documentation Assessment	- Sensitivity Analyses
- Requirements Trace	- Face Validation
- Methodology Review	- Benchmarking
- Code Walkthrough	- Test / Field Data Comparison
- Data Certification...	- Peer / Red-Team Review...

V&V Activity Classes

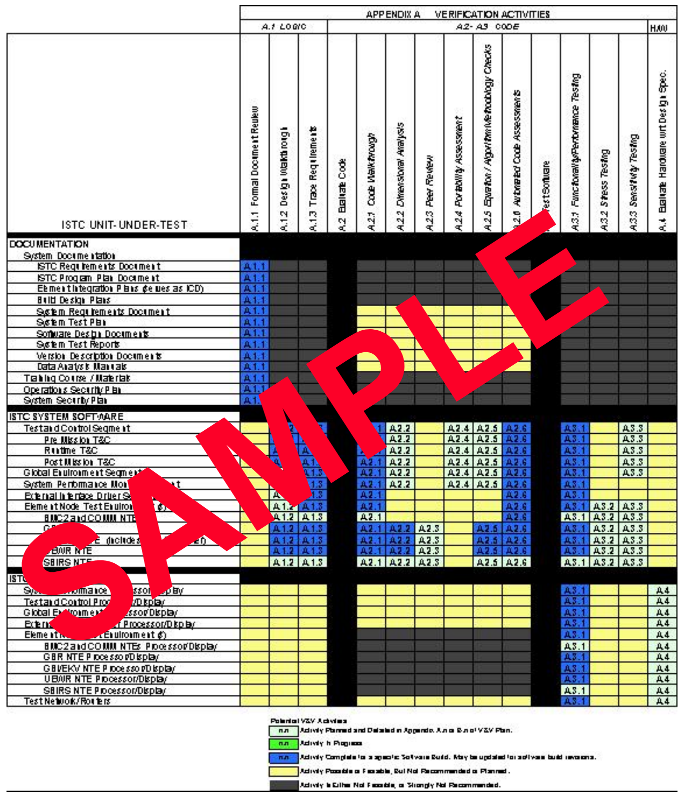
A wide variety of agents are available to most M&S Program Managers. Each V&V agent should be selected based on their capability to serve as the appropriate executor of one of more activities. Each agent's role must be clearly defined. These agents can contribute to the execution of V&V activities that and bring special capabilities to their respective efforts.

For example, for the government's Missile Defense Integrated System Test Capability (ISTC) hardware-in-the-loop simulation VV&A program, the ISTC PM was responsible for overall program strategy and oversight. AEgis Technologies served as the Lead V&V Agent (contractor) and conducted a wide range of independent verification and validation activities for the ISTC. Collectively, the V&V contractor staff was expected to conduct documentation reviews, code reviews, and independent software tests; provide SME support for simulation-to-simulation comparisons; simulation to flight test correlations, and conduct peer reviews and hands-on evaluations.

In addition, the ISTC developer provided systematic ISTC product development, executed selected system, software, or model verification, testing, and prepared system and software documentation. A SETA contractor was directed to conduct ISTC system and design document reviews to assess the maturity of the user documentation. Other government agencies and their support contractor organizations provided SME as requested for reviews and VV&A analysis. The OTA also contributed to accreditation planning efforts, elected to conduct some independent data reviews, and provided some SME support. Coordination among this diverse set of V&V agents was required to execute a balanced, comprehensive VV&A program for ISTC which ultimately led to its formal accreditation by the Ballistic Missile Defense Organization [now the Missile Defense Agency (MDA)], the Air Force Operational Test and Evaluation Center (AFOTEC), and the Army Test and Evaluation Command (ATEC) for a DOD-level Deployment Readiness Review for National Missile Defense.

Planning VV&A Activities

As the Lead V&V Agent, AEgis Technologies was responsible for planning and coordinating the overall VV&A program for ISTC. For the ISTC, AEgis developed a crosswalk of **Evaluation Activities** versus **UUT**, using spreadsheets generated with commercial, off-the-shelf (COTS) software; using this spreadsheet software the cells traced to existing Technical Assessment Procedures (TAP) that identified executing **procedures** and the expected **assessment data products**. This process supported the exploration of alternative suites of V&V program activities that were correlated to low, medium, and high risk and high, medium, and low investment programs (see Figure).



Sample UUT – Activity Matrix

This representation of the evaluation space (i.e., the space projected onto the Evaluation Activities-UUT plane) was a valuable representation of the V&V activity. It also was a convenient medium to support the balancing of 'investment' in V&V and test activities, and it was a simple form from which to generate the cost and resource estimates required to execute the proposed program. Such a representation assured a systematic, complete (but not exhaustive) basis for describing and revising the proposed V&V program, and further tailoring as required during the process of staffing and obtaining approval of the ISTC VV&A Plan.

In summary, the M&S VV&A practitioner who refuses to face the fact that exhaustive V&V is impractical chooses to seek an impossible level of testing. This is a purely political refuge. When Management ships the product over the inevitable objections, the test manager can blame Management for every bug found in the field: “I told them it needed more testing!” Within the testing industry, we’re really struggling with how to know when to say when. For the last few years we have been discussing and debating it. We are meeting and working with a growing community of testers, test managers, and consultants who consider themselves “Good Enough” proponents, and the time is right to propose specific models of “Good Enough” testing.

Risk Based Approaches – JASA

The Joint Accreditation Support Activity (JASA) has developed and documented an approach to VV&A for models and simulations which is based on an assessment of application risks. (Here by "application" is meant the intended use of the M&S, not the M&S software itself.) This approach looks not only at verification and validation activities, but also at all of the activities that provide information pertinent to determining the credibility of M&S, and whether that credibility is sufficient to support its use for a particular application. The approach has been documented in a number of articles and publications, including the proceedings of the 2000 Summer Computer Simulation Conference[13].

Key Factors in Simulation Credibility

JASA has worked extensively with "legacy" M&S that were developed before there was much community emphasis on VV&A. Often it is the case that for those types of simulations, little documentation is available on the original conceptual model or the detailed design. Without that type of information, verification is not really possible, and trying to reverse engineer the simulation after the fact is very expensive, time consuming and only partially successful (often it is difficult to intuit why someone chose the approach they did or where they got embedded data or algorithms). Also, because of the complexity and cost of M&S validation (comparison with 'real world' test data), much of the JASA approach is centered on determining when a user can rely on other measures of simulation credibility that are less costly than traditional V&V. What are those measures, and what kind of confidence do they contribute to simulation outputs? There are five key factors that JASA evaluates to measure simulation credibility:

Capability

Simulations are abstractions of reality; they do not simulate all aspects of the "real world", nor do they need to. In order to be considered credible for use in a particular application, a simulation need only represent those aspects of the "real world" that are important to the intended use. These "capability requirements" are derived from an analysis of the application in which the simulation will be used. These requirements must then be compared to actual simulation capabilities to determine whether or not the simulation has all the features necessary to produce credible outputs for the intended use.

Descriptions of simulation capability (and requirements) should include:

- A clear description of the purpose for which the simulation was developed;

[13] *An Integrated Approach to Evaluating Simulation Credibility*, by Dr. Paul Muessig, Dennis Laack and John Wrobleski, Society for Computer Simulation

- A listing of the physical entities included in the simulation, the functions they perform within the simulation, and the degree of fidelity to which these functions are simulated (conceptual model);
- A description of the environment in which the physical entities interact within the simulation, and the rules under which different entities interact with each other and with the environment (architecture);
- A summary of assumptions and limitations in simulation design and implementation that impact the scope of potential applications for which the simulation can be credibly used.

The credibility of a simulation is a function of its ability to meet the most important elements of required capability as determined by the intended use. While intuitively obvious in principle, in practice both simulation capability requirements and descriptive information about simulations are rarely documented in terms that allow for easy comparison and evaluation.

Software Accuracy

Software accuracy is the degree of 'error-freeness' of the simulation software. One must be able to demonstrate on the basis of software test results not only that the software passed all the planned qualification tests, but that the nature, scope and depth of those tests was matched to the complexity of the simulation and the risks associated with simulation use.

Another factor of equal importance to software test results is the quality of the resources applied to software testing. The more complex a simulation is, the more important it is to apply seasoned resources to the planning, management and execution of the test program. It is not enough to rely on "best commercial practices"; one must establish that these practices are actually being applied by qualified personnel. This includes the software development process used, the quality and appropriateness of the resources applied, and the nature, scope and depth of the artifacts produced. Resources here include the qualifications of the people conducting the tasks, the work-hours applied, and the availability of supporting tools.

Data Accuracy

Data accuracy includes: (1) the appropriateness and error-freeness of all simulation data; and (2) the accuracy of any data transformations performed to convert data from one form to another. This applies to both embedded data and run-time (input) data. Embedded data accuracy is normally addressed during software development and testing; however documentation of embedded data accuracy is often neglected in typical software development documentation. Run-time data accuracy tends to be handled rather passively: as long as the data are obtained from a recognized source, the assumption is made that the data are both appropriate and accurate, often without further inspection.

Data accuracy is also a function of the accuracy of data transformations within the simulation. Unit conversion, coordinate transformations, and data pre- or post-processing algorithms all need to be tested to ensure that good data going in do not become corrupted before being acted upon by simulation algorithms. The nature of these activities tends to be informal, however, and their scope and depth need to be planned against the risk associated with the intended use of the simulation.

Results Accuracy

Results accuracy is the degree of correlation between simulation predictions and real world observations. This is where the term "validation" applies, but there are different types of validation, each of which carries implicit assumptions about how the "real world" is defined:

- Validation against Other Simulations. Comparison of simulation outputs with the outputs of other "accepted" simulations is called "benchmarking". The value of the comparison depends on how credible the "accepted" simulation is.
- Validation against Expert Opinion. Here, simulation design and outputs resulting from well-defined input conditions are reviewed and evaluated by SME. This process is usually called "face validation". The value of face validation as a contributor to simulation credibility depends upon the nature, scope and depth of SME experience relative to the type of simulation being evaluated.
- Validation against Test Data. The value of this method of assessing simulation credibility depends on the credibility of the data used to compare with simulation outputs (test instrumentation used, its inherent measurement accuracy, its calibration history, and any other characteristics that might impact the validity of the validation data set).

Usability

Simulations are credible only within a well-defined usage context, and only when they are properly used within that context. Any simulation attribute that reduces the probability of simulation misuse enhances its credibility within a given context. By usability we mean that collection of simulation user support features that facilitate credible use of the simulation and reduce the probability that it will be employed inappropriately. Examples of such features are: availability of training in proper simulation use and interpretation of outputs; accurate and comprehensive simulation documentation (User's Manuals, Analyst Manuals, Programmer's Manuals, etc.); on-call technical support for simulation users; simulation user groups that meet on a regular basis; the existence and implementation of a sound configuration management process for the simulation, both during and after development; the availability of trained simulation operators and analysts who can run the simulation and interpret its outputs correctly; and any other support feature that can help simulation users ensure credible use of the simulation.

Resources Required Depend on Risk

The resources required (in the 5 categories described above) to determine if a simulation is credible are directly related to how much information is needed to establish that credibility for a particular application. How much information is required depends in turn on the level of risk associated with using the simulation. If you have a high risk application, you're going to need to spend more resources on establishing credibility than if you have a low risk application (where a wrong answer has fewer adverse consequences).

Risk is a concept easily understood but difficult to quantify in objective terms (as discussed previously). Even so, several texts and documents have attempted to introduce some degree of uniformity in quantifying risk.[14] As we discussed early in this paper, one very common approach is to consider risk as the product of two components: the impact (or consequences) of an event and the probability or frequency of the event's occurrence. In most cases the factors in this "equation" cannot be quantified absolutely, but can be subjectively evaluated using principles similar to those described in MIL-STD-882C, "System Safety Program Requirements".

The general process that JASA employs for quantifying risk first puts values on the impact severity and probability for each separately identified risk factor.[15] Using these two values an overall level of risk is assigned to each risk factor (death, damage, delay, etc.). The highest level of risk associated with any risk factor (usually identified as low, moderate, high risk, etc.) is selected as the level that drives the simulation credibility requirement. The criteria used in each step of the risk assessment process are all subjective, but they are explicitly stated, subject to expert review and consensus, and can be tailored to the specifics of individual problems. The details of this process have been described by Muessig et al.[16]

Once the risk factors have been evaluated, the user has some quantifiable (albeit subjective) assessment of the risk associated with using simulation outputs to support the application. Basically, risk assessment shows what bad things can happen if the simulation outputs are wrong but they are used anyway; it also shows how bad those things are (impact severity) and the likelihood of their happening (probability). Based on that risk level, the amount and type of information needed to make an adequate assessment of simulation credibility can be determined (information about Capability, Accuracy and Usability). This is done using the Accreditation Information Requirements Guide (AIRGuide) that has been developed by JASA.

Accreditation Information Requirements Guide

The guide is divided into six major sections (see the accompanying Tables). The first section addresses simulation credibility requirements (that is, the requirements that are determined by the accreditation authority); the other five sections address the five key elements of simulation credibility discussed above (capability, three types of accuracy, and usability). The first column of each table identifies the major questions associated with each of the credibility components. The next column of each table identifies the type(s) of information that may be used to answer

---

14 See, for example, Steele, Lowell W. 1989. *Managing Technology, The Strategic View*. McGraw-Hill, pg 118

15 A risk factor is a specific type of outcome or result. For example, one risk factor might be injury or death of personnel; another might be damage to equipment; a third might be damage to a particular part of the environment.

16 Muessig, P. R., Laack, D. R. and Wrobleski, J. J. "Optimizing the Selection of VV&A Activities: A Risk/Benefit Approach", *Proceedings of the 1997 Summer Computer Simulation Conference*, Arlington VA, pp 855-860

each of the questions. In many cases, there are several types of information that apply to a single question. The third column identifies specific sources for each information type. These three columns basically define the information space that establishes simulation credibility.

The three columns on the far right side of each table provide guidance as to what information is needed to mitigate each level of risk. Note that greater levels of risk require more detailed information to establish simulation credibility. Note also that the assignment of specific information requirements to specific levels of risk is subjective, and should be tailored to meet the requirements of individual applications. The assignments listed in the tables are typical. In some cases, the table provides some flexibility to allow the user to select from two or three alternative information sources to establish the required level of credibility.

For example, in the M&S Credibility Requirements Matrix, the first issue is, "What do you need the simulation to do?" In order to answer that question, you will need an application description and a set of M&S requirements derived from that description (shown in the second column). The third column of the matrix contains a brief description of an Application Description and the resulting M&S requirements. The fourth column in the matrix describes typical sources for the information that would be part of an Application Description and M&S requirements, or where you might find the necessary information to derive them. And the last three columns of the matrix give typical guidelines for whether these information elements are required, and how formally, by application risk (high, moderate or low): for low risk applications of the M&S, the table assumes that verbal descriptions of the application and M&S requirements are adequate, but for moderate and high risk applications these should be documented more formally. The entries in the last three columns of these matrices can be tailored to the specific requirements of the M&S user.

M&S Credibility Requirements Matrix

M&S Credibility Requirements Issue	Items Required	Item Description	Typical Sources	Needed When Risk Is...		
				Low	Moderate	High
What do you need the simulation to do?	Application Description and M&S Requirements	The Application Description defines the overall problem to be solved, the outputs needed from the simulation, and how these outputs will be used in its solution. Specifically describes what the model will be used for and how it will be used in the context of the problem. M&S Requirements must include functional, fidelity and operating requirements. Also specifies required input data sets (e.g., embedded system parameters and constants, look-up tables, run-time inputs, etc.) and input data quality (accuracy) requirements.	Application Description must be developed for each specific application based on what the simulation will be used for. If the simulation is being developed for the specific application, the Application Description and some of the M&S Requirements can sometimes be derived (or inferred) from top level S/W Design documentation. Some programs actually develop an "intended use statement" that identifies some of the items mentioned, but it is rarely comprehensive enough by itself to meet the requirements of this information element. Some of the M&S Requirements are unique to each applciation and must be generated by the analysts who will be using the M&S either via documented study requirements or interviews.	Verbal Description is Sufficient	Req'd	Req'd
How much confidence do you need in simulation outputs?	Risk Analysis Results	Identifies type of risks arising from potential errors in simulation outputs, assesses the probability of the risk actually occurring, and determines the impact of the risk. Result is an assessment of the level of risk associated with the application, which sets the scope and depth of V&V (and related) activities required to mitigate this level of risk.	Must be developed for each specific application. It is essential to obtain consensus on risk elements, impacts, probabilities and levels among technical, operational, and management personnel. These people must be intimately familiar with the decisions to be made based on simulation outputs, as well as the system(s) being simulated and their use. The risk analysis should be based on the Application Description.	Informal	Formal with Documented Results	Formal with Documented Results

M&S Capability Requirements Matrix

<table>
<tr><th>M&S Capability Issue</th><th>Items Required</th><th>Item Description</th><th>Typical Sources</th><th colspan="3">Needed When Risk Is...</th></tr>
<tr><th></th><th></th><th></th><th></th><th>Low</th><th>Moderate</th><th>High</th></tr>
<tr><td rowspan="3">Does the simulation do what you need it to do?</td><td rowspan="3">Functional Breakdown and Description of Model</td><td rowspan="3">Describes what the model actually does. Must describe: M&S functions and relationships between functions; the level of fidelity at which each function is modeled; function level I/O and I/O relationships between functions; the hardware, software, and training needed to operate the model properly and interpret the outputs correctly.</td><td>User Documentation (Users' Manual, Programmers' Manual, Analysts' Manual)</td><td rowspan="3">Any One</td><td>Req'd</td><td>Req'd</td></tr>
<tr><td>Software Design Documentation, possibly including Data Flow Diagrams</td><td rowspan="3">Any One</td><td rowspan="2">Either One</td></tr>
<tr><td>Conceptual Model documentation.</td></tr>
<tr><td></td><td>List of Limitations due to Assumptions and Errors</td><td>Describes model assumptions and known errors, and assesses their impact on model use. The resulting limitations should be correlated with each of the functions in the Functional Breakdown, but may also be useful at the overall simulation level. List should ID any and all assumptions and/or errors of each M&S function (or of the model as a whole) that are implicit or explicit in the model's design and/or coding, as well as the implications of these limitations on appropriate or acceptable uses of the simulation.</td><td>Software Design Documentation, User Documentation are the most typical sources for inherent assumptions and limitations arising from the algorithms used. Configuration Management Databases are useful sources for known errors. Some assumptions and limitations may also be found in Verification or Validation Reports but may not be explicitly stated as an assumption, limitation, or error.</td><td>Desirable</td><td>Req'd</td></tr>
</table>

M&S Usability Requirements Matrix

M&S Usability Issue	Items Required	Item Description	Typical Sources	Needed When Risk Is...		
				Low	Moderate	High
Do you have confidence that the simulation is being run properly?	Demonstration of the Computer Hardware and operating system suitability	Test results that show that the hardware and operating systems used to host the model or simulation (if different than that used to develop the M&S) will allow it to run correctly and produce consistent results across platforms.	Information on M&S portability across platforms is usually found in the user documentation associated with the simulation. If this information is not documented, test results will be needed to demonstrate portability.	Req'd	Req'd	Req'd
	Operator Qualifications	Information to demonstrate that the operators running the simulation have the expertise and knowledge to properly set up the simulation, execute it, and operate all associated tools and utilities. Typical information includes experience with the specific model being used, formal training on the model, and experience with the hardware, software, and interface devices being used.	This information is usually gathered from biographies or interviews with the operators.	Req'd	Req'd	Req'd
Can you convince others of your interpretation of simulation outputs?	Analyst Qualifications	Information to demonstrate that the analysts using the simulation have the expertise and knowledge to properly generate the input data and interpret the outputs. Typical information includes experience with the specific model being used, formal training on the model, experience in performing similar analyses and experience or training in M&S based analysis techniques.	This information is usually gathered from biographies or interviews with the analysts. It may also be found in prior accreditation assessment reports.	Req'd	Req'd	Req'd
	Demonstration of Pre- and Post- processor acceptability	Information that shows that any auxiliary tools and utilities used to format or load input data, or to convert, record, and visualize model outputs are suitable for the intended purpose(s). The type of information usually presented includes descriptive documentation of the tools and utilities being used for these purposes.	User documentation associated with the simulation may list tools and utilities that are compatible with it. If this information is lacking, user documentation for the tools and utilities may contain information that will aid the determination of tool compatibilty with the simulation.			Req'd

M&S Software Accuracy Requirements Matrix

M&S Software Accuracy Issue: How much confidence do you have in the accuracy of the software?

Items Required	**Item Description**	**Typical Sources**	**Needed When Risk is …**		
			Low	**Medium**	**High**
S/W Development Process Description	The process description should include a description of the development paradigm and how it is being implemented (including the use of CASE tools); a logical process for defining, tracing, and testing requirements throughout development; configuration management during the development process; and adequate provision for documentation of all of these activities	Look for a S/W Development Plan (SDP) or a Configuration Management Plan that outlines the development process used. If the development is underway, these plans should describe the process currently being used.	Either One	Required	Required

Items Required	Item Description	Typical Sources	Needed When Risk is …		
			Low	**Medium**	**High**
S/W Development Resources Description	The resource description should include a description of the H/W environment and the S/W engineering tools that will be/were used; the qualifications of the personnel who will/did code the S/W and perform CM functions; and who will be/was responsible for production of key documentation and testing. A history of similar simulation development experience should also be included.	Check the SDP or other management plans that might contain such information. IF this information isn't in existing documentation, discussion with the software developers and managers will be necessary to obtain as much of this information as possible, even if anecdotal. Evidence of simulation development qualifications may be available in SEI Capability Maturity Model evaluation reports.		Any two	Required

Items Required	Item Description	Typical Sources	Needed When Risk is …		
			Low	**Medium**	**High**
S/W Development Artifacts	"Artifacts" refers to the evidence (usually documentary in nature) that S/W development is actually being implemented in accordance with the guidelines and specifications called out in the SDP (or its equivalent). Documentary artifacts must also be in compliance with known (or acceptable) standards and practices for format, content, currency and applicability to the current version of the S/W.	Look for standard documentation that reflects the current state of the S/W and that conforms to known standards of information content. The most important examples are configuration management histories and logs. Model Documentation (User Manual, Programmers' Manual, etc.) SW Design documentation (particularly a documented set of requirements and a conceptual model)			Any Two
S/W development Results	V&V results include all evidence that the code has been developed according to the design and are free of critical errors.	Requirements Trace Reports Reports of design Reviews, Peer Reviews, and/or Logical Reviews	Either One	Either One	Either One

Items Required	Item Description	Typical Sources	Needed When Risk is …		
			Low	**Medium**	**High**
	The types of results will include reports from design reviews, code walk-throughs, regression testing on model changes, S/W testing, and supplemental V&V efforts of previous M&S users.	Code Walkthrough Reports S/W Problem Change Request Logs Module S/W Test Reports Subsystem S/W Test Reports System S/W Test Reports	Any One	Any Two	Any Three
S/W Management Process Description	The process description should include a description of the post development management of the software. This should include processes for documenting, implementing, tracking and testing M&S changes resulting from either requirements changes or software errors. Processes should also exist for keeping all software documentation current with the software.	Check the S/W Management Plan, Configuration Management Plan, V&V Plan, etc. to determine if they address M&S life cycle activities. If these documents only address development activities, look for other documentation that describes the life-cycle management activities	Any Two	Required	Required

Items Required	Item Description	Typical Sources	Needed When Risk is …		
			Low	Medium	High
S/W Management Resources Description	The resource description should summarize the nature and extent of resources currently being applied to simulation management and support. The information should indicate whether sufficient funding and experienced personnel are being applied to ongoing documentation support, configuration management support, regression testing, user group support, training, technical support, etc.	Check any management plans that might contain such information. If this information is not in existing documentation, discussion with the model managers and/or software developers will be necessary to obtain as much of this information as possible, even if anecdotal.		Desired	Required
S/W Management Artifacts	Artifacts refer to the evidence (usually documentary in nature) that S/W maintenance is actually being conducted in accordance with the guidelines and specifications called out in the SMP or its equivalent.	Configuration Management Database status reports, system change requests (SCR) and/or system trouble reports Up to date model documentation (User Manual, Programmers' Manual, etc.) CCB and User Group meeting minutes Updated S/W Design documentation		Any One	Any Two

Items Required	**Item Description**	**Typical Sources**	**Needed When Risk is …**		
			Low	**Medium**	**High**
Post development S/W V&V Results		S/W Problem Change Request (SPCR) logs that correlate V&V results with specific versions of the software Alpha- or beta- test reports for both new requirements testing and regression testing		Either one coupled with S/S mgmt items above	Either one coupled with S/S mgmt items above
		Specific verification reports for the M&S version being used	Note 1	Note 1	Note 1
		History of successful usage in similar applications	Note 2		
Note 1: If the scope and depth of the specific verification results equates to the scope and depth of development V&V required for a particular level of risk (Item 4D), this element can be substituted for all the above items dealing with the quality of the software. Note 2: This item alone can be used as evidence of sufficient quality for low risk applications (it can replace all items above for Issue 4)					

M&S Data Accuracy Requirements Matrix

M&S Data Accuracy Issue	Items Required	Item Description	Typical Sources	Needed When Risk Is...		
				Low	Moderate	High
How much confidence do you have in the quality and suitability of input data obtained from outside sources?	Data Quality Profile	A body of metadata that describes the database, its source, specifications, intended usage, history, and how it was collected. Metadata elements should exist at the database, data element, and data value levels.	Metadata elements should be available from the data producer or may exist in the same archives that contain the database itself. A list of essential metadata elements considered necessary for different risk levels is contained in Sheet 2 of this Excel Workbook	Either one. Depth of Information indicated in Data Accuracy Sheet 2	Either one. Depth of Information indicated in Data Accuracy Sheet 2	Either one. Depth of Information indicated in Data Acuuracy Sheet 2
	Independent Assessment of Data Quality	An independent assessment is prepared by the data user when the Data Quality Profile is inadequate, incomplete, or does not exist. This assessment addresses the key metadata elements listed in the Data Quality Profile.	Information that indicates the quality of test data can generally be found in documents such as test plans, laboratory procedures, calibration records, test reports, etc. Information that indicates the quality of data collected through surveys or monitoring operations can generally be found in data collection plans, reports, and raw notes.			
	Data Manipulation Verification	This item refers to the verification of any data manipulation done by the user. "Data manipulation" includes operations such as: editing, subset selection, merging, aggregation, transformation, estimation, interpolation, etc. Verification includes any activities that are done to ensure that the data manipulation steps are correct and do not introduce unknown errors.	Verfication of data manipulation procedures may be documented in M&S verification reports (when done in conjunction with M&S development). Other data manipulation should be reviewed and verified as part of the M&S accreditaiton process and documented in the accreditation report. This documentation should describe the verification techniques that were used.	Req'd at cursory level	Req'd	Req'd
How much confidence do you have in the quality and suitability of self-generated input data ?	Description of Data Quality Assurance Process for self-generated data	An assessment by the data user of the process, equipment, tools, instrumentation, etc. used in generating the data. This assessment should generate information similar to that included in the critical metadata elements of the Data Quality Profile.	Information that indicates the quality of test data can generally be found in documents such as test plans, laboratory procedures, calibration records, test reports, etc. Information that indicates the quality of data collected through surveys or monitoring operations can generally be found in data collection plans, reports, and raw notes.	Req'd. Depth of Information indicated in sheet 2	Req'd. Depth of Information indicated in sheet 2	Req'd. Depth of Information indicated in sheet 2
	Description of Data Quality Assurance Resources for self-generated data	This item refers to the verification of any data manipulation done by the user. "Data manipulation" includes operations such as: editing, subset selection, merging, aggregation, transformation, estimation, interpolation, etc. Verification includes any activities that are done to ensure that the data manipulation steps are correct and do not introduce unknown errors.	Verfication of data manipulation procedures may be documented in M&S verification reports (when done in conjunction with M&S development). Other data manipulation should be reviewed and verified as part of the M&S accreditaiton process and documented in the accreditation report. This documentation should describe the verification techniques that were used.	Req'd at cursory level	Req'd	Req'd

M&S Output Accuracy Requirements Matrix

<table>
<tr><th rowspan="2">M&S Output Accuracy Issue</th><th rowspan="2">Items Required</th><th rowspan="2">Item Description</th><th rowspan="2">Typical Sources</th><th colspan="3">Needed When Risk Is...</th></tr>
<tr><th>Low</th><th>Moderate</th><th>High</th></tr>
<tr><td>How much confidence do you have in the simulation outputs?</td><td>Benchmarking Results</td><td>Documents the results of comparisons between simulation (or simulation component) outputs and those of a "standard" or widely accepted simulation. Benchmark simulations are generally of greater fidelity than the simulation (or component) under review and are characterized by some "stamp of approval" from a recognized authority. Benchmark results should include the name and source of the standard simulation, why it is (or should be) considered a "reference" simulation, which parameters between simulations (or components) were compared (and why), what the results of the comparison were, and what these results imply about the credibility of the outputs from the simulation under review.</td><td>Benchmarking results are usually documented in either a validation report, a briefing that describes the results of the comparison, or an Accreditation Support Package for certain models. These reports would generally be prepared by previous users. They might also be available through the model manager or individual Service M&S databases. If these results are for a previous model version, they must be accompanied by model management activities that are appropriate to your application risk (as outlined in Items 4E, 4F, and 4G).</td><td rowspan="3">Any One</td><td></td><td rowspan="2">Either both of these</td></tr>
<tr><td></td><td>Face Validation Results</td><td>Describes the results of subject matter expert opinions about simulation realism and accuracy. This should be based on a structured review of simulation (or component) outputs, sensitivities, and/or design. When face validation is a review of the simulation design, the documentation should state whether the representations are realistic and whether any assumptions that underly the design are acceptable from the perspective of the intended use. Documentation should describe which aspects of the simulation were reviewed (and why), who participated in the review, why one should trust their opinions (e.g., biographical information), what the results of the review were, and what these results imply about the credibility of the simulation.</td><td>Face validation results are typically documented in a Face Validation report (or an Accreditation Support Package for certain models) or a previous Accreditation Assessment Report (if the face validation was done as part of an accreditation assessment). If the review was a validation of the design, the results may be reported in a design review report (either a formal report or an annotated briefing). These reports would generally be prepared by previous users. They might also be available through the model manager or individual Service M&S databases. If these results are for a previous model version, they must be accompanied by model management activities that are appropriate to your application risk (as outlined in Items 4E, 4F, and 4G).</td><td rowspan="2">Either One</td></tr>
<tr><td></td><td>"Results" Validation Documentation</td><td>Describes the results of comparisons between simulation (or component) outputs and data collected from tests or operation of the real system(s) or process(es) being simulated. The documentation should include a description of the source data used in the comparison, from where and how it was obtained, and why it should be considered representative of the real world. Issues relating to data quality (e.g. instrumentation accuracy, calibration, test scenario realism, etc.) should be addressed in the validation report. The correlation between model outputs and real world data should be stated in quantitative terms rather than merely stating that the correlation was "good". Any anomalies and their impact on model usage should be explained.</td><td>"Results" validation is typically documented in a Validation Report (or an Accreditation Support Package for certain models). In some cases, results validation might be documented with an annotated briefing. These reports would generally be prepared by the simulation developer or previous users. They might also be available through the model manager or individual Service M&S databases. If these results are for a previous model version, they must be accompanied by model management activities that are appropriate to your application risk (as outlined in Items 4E, 4F, and 4G).</td><td>Or this item is required</td></tr>
</table>

In practice, the process works as follows. The risk level is determined by combining impact and probability for various types of outcomes. Then the tables are used to determine how much and what type of information is required to establish the required level of simulation credibility. These information requirements are then compared with available information about the simulation, and a list of credibility 'shortfalls' is compiled. Each element of this list is then evaluated for its impact on application risk. Unmet requirements for simulation credibility that have acceptable (i.e., low risk) work-arounds are removed from the list. Unmet requirements for simulation credibility that have no acceptable work-arounds generate a requirement for more detailed information in the appropriate category. The user will have to provide the resources necessary to generate that information. Alternatively, if the user cannot provide more funding or chooses not to, the user can choose to use the model with the amount of evidence available and accept a higher level of risk.

These tables are only examples of the process. Simulation users should tailor this process to the particular circumstances of their specific applications. A factor that should be included in tailoring the process is the fact that in practice, model users (and developers) may have agency or industry policies about model credibility or VV&A with which they must comply. For example, within DOD each Service has M&S VV&A policies, and some agencies within the Services have their own policy. The challenge in coming up with workable VV&A plans for DOD acquisition programs has been to figure out how to do something meaningful that complies with policy and that can really be done within the time, funding, expertise, and validation data available. The concepts in the accreditation information requirements guide have helped to tailor the policy guidance, but from a political point of view, the customer has been most interested in showing he complied with Service policy. Program Managers are interested in whether the evidence of simulation credibility is sufficient to justify confidence in the model for the intended purpose. But in justifying their accreditation case to outsiders, compliance with policy was a better story than the fact that the case was adequate and compelling.

Resource Requirements

The benefit of categorizing the key elements of simulation credibility as described above is that it provides a convenient way to associate standard V&V activities with the types of credibility they provide. The categories of simulation credibility also serve to point out areas where standard V&V activities fall short of fully addressing all aspects of simulation credibility, and they suggest other types of information that might be equally important. The result is a set of metrics by which simulation credibility can be evaluated. The risk techniques outlined here allow simulation users to determine how much and what specific types of information are needed to establish sufficient credibility for their intended application. This information can form a convenient basis both for V&V and simulation credibility assessment planning.

Because the risk-based approach determines what V&V activities are required for a specific application, if a 'rule of thumb' can be derived for what resources are required for each activity, then we can determine costs associated with plans for V&V and credibility assessment. Later in this paper we will see that the Susceptibility Model Assessment with Range Test (SMART) program developed estimates of the resources required to perform many of these standard V&V

activities (this was for legacy models, well after the model was already developed). By combining this risk-based approach to determine what activities are needed with the experience of that program on what it costs to accomplish those activities, we can estimate the costs of V&V as a function of risk

For example, the table below lists the activities necessary to describe the capabilities of the simulation. In that table, there are two basic elements: the "Functional Breakdown and Description of the Model", and 'List of Limitations Due to Assumptions and Errors". Those two activities were accomplished under the SMART project, which experienced average resource requirements of 2 work-months to develop the functional breakdown of the model, plus 2 work-months for a detailed description of each "functional element" within the model. To develop the list of assumptions, limitations and errors took 3 work-months, on average.

These figures were averaged over five engineering level M&S for work done essentially to re-engineer them. The re-engineering was done by subject matter experts, some of whom worked on the original M&S development. A detailed description of what each task entailed as well as a standard reporting format were available, so little time was wasted trying to figure out what the task really was and how to report on the results.

These cost figures taken from re-engineering efforts may or may not be relevant to what it would cost to do these tasks as part of development (for example to develop the conceptual model at the beginning of the M&S development project). However, because of the level of expertise available to the SMART program for each model (in most cases the original model developer was hired to develop the conceptual model after the fact), these levels of effort should be somewhat representative of any M&S project.

Using the V&V task breakdown suggested by the table, low-risk applications require a functional breakdown of the model for its capability description, which will take approximately 2 work-months. For a moderate risk application, an assessment of assumptions, limitations and errors is required in addition to that, for another 3 work-months. And for high risk applications, a detailed conceptual model will be required in addition, which will take approximately 2 work-months per major simulation functional element. A further assessment can be made to determine which of the model's functional elements are critical to that application, or have the greatest impact on simulation outputs, in order to reduce the overall number of conceptual model elements that are absolutely required.

The table summarizes this example.

Example Resource Requirements for "M&S Capability Description" (In work-months)

Information Element	Application Risk Level		
	Low	Moderate	High
Functional Breakdown	2	2	2
Functional Element Description (conceptual model)			2/FE
Summary of Limitations due to assumptions and errors		3	3
Total Resource Requirement (work-months)	**2**	**5**	**5+2/FE**

Risk Based Approaches – U.K.

In a paper entitled "Verification, Validation and Accreditation of Models and Simulations Used for Test and Evaluation – A Risk/Benefit Approach" in March, 1999[17] the U.K Ministry of Defense (DERA) described a risk-based approach to planning V&V efforts. In a similar fashion to the JASA approach described above, the U.K paper describes a formal hazard/risk/benefit analysis process, and based on the results of that analysis they select appropriate verification and validation activities. The aim of their process is to identify the hazards associated with use of the M&S, and to take the necessary steps (V&V and other activities) to reduce the likelihood or consequences of these hazards. The ultimate aim of the approach is to reduce the V&V activities that are required to a minimum set determined by the risks associated with the application. They also include a provision for a benefits analysis, in which case it can be demonstrated that V&V activities can also increase the likelihood of realising the benefits of M&S use. The author of this paper originally came from a safety background and he seemed to find that many of the fundamental principles of safety analysis transferred easily over to M&S VV&A.

The UK approach is divided into several distinct steps. The first produces a complete list of the functions required by the M&S to meet the specified purpose. Once this list of required functions (or capabilities) has been developed, a functional failure analysis can be completed. This consists of developing a table containing the complete list of M&S functions, their failure modes, the domains in which their impact is to be assessed, the estimated most critical phase/condition, the effect of the functional failure at the system level, and the estimated maximum possible severity category in each of the impact domains. This basically constitutes a hazard analysis of the use of M&S for the application. The results of this analysis provide the same table of impact domains and severities that is developed in the JASA approach; the concepts behind both tables (U.K. and JASA) come from the safety analysis community.

A further refinement of the general risk-based approach is provided by the U.K. paper. Consequence Analysis is a diagrammatic technique which can be used to advantage when more than one impact domain is to be addressed during the analysis. Using this technique, each hazard requiring further analysis becomes the trigger in a series of events, passing through various decision points until the consequences at the system level can be determined. Taking different branches at each of the decision points can result in different consequences.

Another refinement provided by the U.K. approach is a "Benefits Analysis". It may be the case that the risks involved with a particular M&S are fairly low, but the potential benefits to be realised are high. It can be assumed that if M&S are being proposed in lieu of "traditional" testing it is expected to bring some form of benefit. In some cases a separate benefits analysis may be deemed appropriate. The benefits analysis is similar to the risk/hazard analysis in approach and concept; benefits can be viewed as positive hazards.

[17] Verification, Validation and Accreditation of Models and Simulations Used for Test and Evaluation: A Risk/Benefit Based Approach, Chris Mugridge, Defence Evaluation and Research Agency, UK, March 1999

Once the levels of risk have been identified, a list of development, verification and validation activities can be selected from the tasks discussed below.

Verification is described, for the purposes of the U.K. approach, in terms of a required *level.* The levels range from Level A (the least stringent) to level E (the most stringent). The individual verification activities which are necessary in order to claim compliance with each of these levels are selected from: Static Analysis, Control Flow Analysis, Data Use Analysis, Information Flow Analysis, Semantic (Functional) Analysis, Dynamic Analysis, Unit Testing, Integration Testing, Conceptual Model Verification, Compiler Verification, Structured Walkthroughs, and Fagan Inspections.

As with verification, validation is described, for the purposes of the U.K. approach, in terms of a required *level.* The levels range from Level 1 (the least stringent) to level 5 (the most stringent). The minimum acceptable level of validation will depend upon the required level of confidence (or the level of acceptable risk). Level 5 is defined as validation with data from field trials; Level 4 as validation with data from hardware-in-the-loop trials; Level 3 as validation with data from components of the system; Level 2 as validation with data from other validated models; and Level 1 as validation with lower levels of data.

Resource requirements for each of the available activities have not been identified. However, the list of available V&V activities (and others listed in the U.K. paper) can be correlated to the historical SMART cost data in the same manner as the JASA approach. Thus the U.K. approach and the JASA approach, using risk-based analyses similar to those developed by the safety community and historical data on V&V activity costs, can both support the development of resource estimating methods for V&V.

An Overview of the Verification and Validation Cost Estimating Tool (CET)

This section addresses the development and use of the Verification and Validation CET, developed by Robert O. Lewis of the Boeing Company. The tool is specific to VV&A cost estimating for all types of M&S including High-Level Architecture (HLA)-based federations. It is a parametric cost model that considers the unique features and characteristics of each project and its intended application, its inherent complexity, the effects of leveraging and risk and uncertainty, and then uses size of the software as the major variable, except for federations (Federations use a percentage of the development cost as the primary variable.) The tool is able to handle unusual or extreme V&V cases without difficulty. It handles both phase-specific and non-phase-specific other direct costs (ODC) in a logical manner using an innovative algorithm and built-in spreadsheet. The tool runs under ™Windows 2000 and requires no special software.

Introduction

The CET has evolved over the past several years with small increments of funding from the TRADOC Analysis Center (TRAC) at Ft. Leavenworth, Kansas, the Army AMIP program, and the Developmental Test Command (DTC) Headquarters, Aberdeen, MD. The development began in the fall of 1998, and the CET is currently a mature product. The current version is directed only at VV&A projects, although it can be used for IV&V efforts as well with the substitution of the appropriate internal model.

Perspective

In the past, cost estimating for VV&A programs has largely been based on coarse estimates of percentages of development cost, but this approach only works for new or recently completed M&S efforts, and even then it is not very accurate. Unique attributes of each program tend to be ignored and each estimate done in this manner is little better than an educated guess. Percentage of development cost can yield especially poor estimates when used on legacy programs that have undergone several or many changes. Should the initial development cost be used? Should each modification be added to the total? Should only the major changes be included? To combine costs without a logical approach could produce a cost estimate two or three times higher than the cost of building a new M&S, so the idea of "percentage of development cost" can, depending on the circumstances, either be a very poor number or a barely acceptable number.

Regardless, a percentage-based estimate is not good enough by itself, because it does not consider size and programming languages, complexity, intended use of the product, risk and uncertainty, unique characteristics of the M&S, amount of reuse, any previous VV&A history, and adequacy of the documentation for starters. Further, it does not consider costs of SME,

™ Windows is a registered Trademark of the MICROSOFT Corporation.

software tools, support software and hardware, communications and networking, travel and TDY, and any other extraordinary costs that may occur.

All of these shortcomings provided the incentives to the developer of the CET to conceive a more scientific and hence better cost estimating approach than has existed previously. The CET must cope equally well with M&S of various types and uses: to be effective, this tool has to support reuse of legacy M&S products with and without good VV&A history, legacy M&S with minor or major modifications, and new stand-alone M&S. Finally, it has to handle distributed simulations, most of which are rapidly becoming HLA-based. In their new RPG, DMSO recognized this same set of problems and tailored their latest products to provide unique solutions for each of these variations in web-based and traditional documentation formats. The cost estimating approach described in this document and the DMSO RPG use essentially the same VV&A process models to ensure consistency and compatibility between the cost estimates and the execution of the VV&A program. The tool also makes re-planning VV&A efforts much quicker and more accurate.

Identifying the Generic Type of M&S and its Application

Cost estimating is a complex process that first has to consider several things about the M&S to which the VV&A is applied. The first dimension of the problem is to identify the type of M&S and something about how it is to be used. The CET evaluates each M&S product (and project) on an individual basis. Incidentally, the V&V effort is decoupled from the Accreditation cost in the CET because the two are estimated independently. The first five examples given below refer to M&S that are simply stand-alone products or possible federates of a federation (federates are individual M&S products that are capable of joining HLA-based federations.) The major M&S types are described as follows:

- The first, simplest, and most economical M&S example is a legacy product that is to be reused essentially "as-is." The concerns here focus immediately on the previous VV&A history and past experience from its users. These VV&A programs are typically low levels of effort, with an occasional part of the M&S that requires more significant effort. The V&V effort can be low as long as the prospective application closely resembles the original applications and the accompanying information, including VV&A history, is complete enough to support the new accreditation.

- The second example is the legacy product that requires minor modifications to make it acceptable for the intended use. In accordance with long-standing configuration management guidelines, we selected 30% change to the software as the cut-off point for defining minor changes. Anything above this amount constitutes a major change or a so-called "heavily modified" legacy product. M&S with minor changes typically require low-to-moderate V&V effort, depending upon which parts changed and how close it pushes the 30% figure. Generally, data changes tend to require lower V&V effort than algorithm changes.

- The third example is the legacy product that undergoes substantial modification (>30%). At this point, most of any prior V&V can no longer be trusted, so the V&V effort is essentially repeated with little reliance on historical data. Nonetheless, there are some savings derived from the unchanged parts that are factored into the cost-estimating algorithm on a proportional basis.

- The fourth type is the new stand-alone M&S. From a VV&A standpoint, this type and the heavily modified legacy product are about the same with a slight adjustment for reused parts in the legacy one. Therefore, one VV&A model works for both examples with allowances for reuse, which can also apply to "new" M&S since there are numerous examples of reuse even in those that claim to be new products.

- The fifth type of M&S is that encompassed by the DTC methodology (the DTC VV&A model has been included at the request of the second sponsor of the CET). It is a comprehensive approach that is similar to those mentioned above, and it too is tailored to the unique attributes of the application. Its inclusion indicates the universality of the cost estimating approach used in the tool and demonstrates the ability of the tool to accept additional models without breakage.

- The sixth type is the distributed simulation. These have undergone several evolutions in the 1990s ranging from the Aggregate Level Simulation Protocol (ALSP), to Distributed Interactive Simulation (DIS), to the current HLA-based products. Most of our distributed simulation emphasis in the cost estimating process is on HLA federations. These federations are composed of several (or many) stand-alone M&S products that become federates when linked together. VV&A of a federation assumes that federates come with a pedigree from previous VV&A. When this is not the case, it must be taken into consideration separately and not be included as part of the federation VV&A cost estimate. Even with careful planning and screening of candidate federates, when they are joined together there is no assurance that they will behave correctly or provide the correct results, so a significant part of the federation development and execution process (called the FEDEP) is intended to see that this occurs. It is not our goal to go back to square one and redo the VV&A of the federates, rather, it is to negotiate with the federate owners or providers to do certain things to ensure their products work as required to support the federation requirements. The VV&A of a federation is mostly focused on how well its overall mission is being met. To do this requires looking at the interactions, behaviors, fidelity, and performance of the elements represented in federates in a total context. This can be so complex in large federations that the roles of specific participants (players) or groups may have to be looked at separately. Thus, the battlefield or mission space will be viewed and analyzed from numerous perspectives and parts of a scenario may work correctly while others can be corrupted or suspect. Because of the complexities and uncertainty of performing VV&A on a federation, the size of individual federates becomes a useless basis of estimate, and percentage of federation cost becomes the best estimating approach. Thus, even the key cost parameters for federations are different from other types of M&S.

Key Cost Estimating Relationships

In the CET VV&A cost estimating process, all M&S except federations use "size of the software product" as the most significant parameter. VV&A estimates for federations, on the other hand, use a percentage of the federation development cost estimate as the most significant parameter. Aside from these very different approaches, other parts of the estimating processes use many of the same process steps and factors to determine which VV&A activities to perform by phase and what their relative intensity should be. In both cases, each activity is assigned a number in the CET referred to as "raw counts" to indicate its relative level of effort. These raw counts are based on a significant historical V&V project sample, surveys from experts, etc. The raw counts are selected uniquely for each type of M&S application and are then tailored by several quality metrics. The user of the cost tool does not have to worry about how the raw counts were derived and assigned or what they mean. Individual V&V activities where leveraging is planned are reduced to a small percentage (10% or less) of the original raw counts; this is anticipated more in federation development than for any other type of product. (Leveraging is the process wherein V&V accepts and credits work done by others, e.g., the developer, as part of the V&V pool of evidence.) Next, both estimating approaches introduce risk and uncertainty to weight the level of effort (now in "adjusted counts") for each activity, but each uses a slightly different set of factors. In both cases, the adjusted counts are totaled by VV&A step (or phase) and for the entire program.

In all cases except federations, the total adjusted counts are applied to an algorithm that converts them to a VV&A cost per line of code for the particular software languages involved, and then multiplies the resulting dollar figure by the number of lines of code. There are several options to help refine the code counts, e.g., Function Points, Logical lines of code, etc. The cost estimating process has to know what the average loaded man-year costs are for the VV&A staff. In any case, the result is the cost of VV&A labor for the entire effort, which, because we know the distribution of counts by activity and phase, can produce their associated budgets in man-hours or man-weeks, etc. Most parametric cost models do the same thing by project phase.

The federation cost estimating process is quite similar up to the point that the process determines the total adjusted counts, which in this case are applied to a different algorithm than the one described above that used VV&A cost per line of code. In this case, the total adjusted counts are used to calculate a VV&A cost based on a percentage of the cost estimate for the federation development. Although the percentage of development cost has long been used as an estimating technique for V&V efforts, no V&V cost estimating process up until now accounts for the following factors in a quantitative manner:

1) Specific task selection based on unique program characteristics,
2) Adjustments based on program particulars such as adequacy of interim products and other quality metrics,
3) Significant allowances for leveraging from work performed by other team members, mostly the developer and sometimes independent analysis and T&E agents, and
4) Weighting by risk and uncertainty.

Cost as a Function of Size or Development Cost

Instead of picking some arbitrary number like 5% of development costs for the VV&A of a reused simulation product or 10% for a new one, the CET estimating process considers the most significant program parameters in generating a cost figure that is much better aligned to the characteristics of the M&S, the needs of the program, and the amount of V&V required for credibility and accreditation.

Before going any further with a discussion on percentages of development cost as a basis of VV&A cost, several points have to be made. For legacy products, the total "sunk" cost can be quite high and may have accumulated over many years and through many revisions. It is, therefore, a poor number to use for estimating VV&A cost. Unless you can identify the cost of individual versions and why they were built and adjust these numbers accordingly (which is usually impossible), there will be serious skewing in the cost estimating process. A much better approach in this case is to use the size of the product that will be reused, identify its software languages (which have a very large effect on cost per line of code), and determine as necessary how much is new versus reused. This technique is even used for new stand-alone M&S, but in that case must be based on development "estimates," not actual counts, as well as other items and issues.

The approach of counting lines of code in estimating VV&A costs for federations makes no sense. Would you count the code in all the federates? No! Would you count the code generated in interfacing all the federates to the federation? Probably yes, but the result would not be a significant factor in the overall estimate and is very difficult to use. This line of questioning can continue, but for federations an answer based on product size always comes up short.

Fortunately, the sponsor and developer (sometimes called the federation integrator) have to make a rather significant planning investment in order to estimate the cost of developing a federation. The development cost estimating process is based on many variables that contribute to the overall understanding of the effort, blended with the experience and systematic estimating process of the development team. In this case, there is intrinsic safety associated with basing the VV&A estimate on the development estimate not found in any totally independent form of estimating. A VV&A estimate tied to the development estimate assumes that if the development scope changes significantly, then VV&A will receive a similar adjustment. This works in either direction, so there are a number of self-regulating aspects to this type of estimate that protect the sponsor's and V&V agent's interests. Since the problem is multi-dimensional, the developer's estimate needs to reflect good knowledge and quantification of the following, which benefit the VV&A planning as well:

- The use case – this example is a new or modified reused HLA federation.
- A clear understanding of the user and sponsor needs and objectives of the federation.
- A clear delineation of roles and responsibilities of each participant in the development effort and how the organizational interfaces will operate (including VV&A).
- A relative idea of the number of federates and their entities and how they are expected to interact in their configuration, including the geographically separated

federates. This helps scope the magnitude and difficulty of the total effort, amount of wide-band communications, interface coordination, etc.

- Specific answers to questions about the federation requirements, conceptual model, input data quality, and stability and appropriateness of the selected federates.
- Estimated LOE or cost for end-to-end execution the FEDEP for this federation.
- Estimated LOE or cost to interface the federates to the HLA RTI and simulation infrastructure (common databases, synthetic environment, networks, etc.)
- Risk and uncertainty of the program and how these factors affect the cost.
- The estimated size of the new software and its relative complexity and fidelity required by the federation (this helps bound the problem and provides sanity checking but is not a primary estimating factor.)
- Measure of federation complexity as measured by the different types of interactions.
- A list of ODC such as tools, facilities, communications, SME, support software, TDY, travel, etc. (The V&V agent needs similar data.)
- Average man-hour, man-month, or man-year cost for the staff. (The V&V agent needs similar data.)

From these factors, the developer can produce a relatively accurate cost proposal for the effort. Because of the scope and comprehensiveness of the developer's cost estimate, the VV&A estimate for federations can then be tied to it as a percentage of that cost. In this cost estimating process, this VV&A percentage is not a fixed number; rather, it is calculated based on program characteristics, amount of reuse, leveraging, risk and uncertainty, etc. in much the same way as for the other types of VV&A.

What Constitutes a Good VV&A Cost Estimate?

Use of terms like "good, adequate, reasonable, and relatively accurate" in VV&A cost estimating means that the activities and tasks outlined for the effort can be accomplished effectively and thoroughly, but not excessively. The term *"should-cost" figure* is very important since it should be produced independently of other inputs, influences, and outside budget constraints. Thus, this figure is what an adequately funded and supported VV&A effort for the particular use case should cost. This should be the initial cost estimate or budget; then, if conditions change during the development of the product, the VV&A budget should be reviewed to ensure it still correctly represents the required effort.

Real-world constraints sometimes affect this VV&A budget. In a typical scenario, someone on the decision-making side of the program decides that the VV&A budget should be a specific dollar amount or percentage of the federation development cost without developing a scientifically-derived budget. Sometimes these numbers are above the should-cost figure, but most of the time they are below. The sad part of this "shoot from the hip approach" is that it has promoted occasional criticism concerning the high costs of VV&A. Conversely, it also results in VV&A efforts that are woefully under-funded, which in turn, fosters criticism that the VV&A effort could not do everything that needed to be done. Either way, VV&A takes the heat for lack of a scientific cost estimating process. The estimating principles and control factors in this tool will not enable "extreme" estimates to be generated in the first place.

Validation of Some M&S Can Present a Serious Estimating Problem

It can be argued that there are simulations that are very difficult to validate for any number of reasons, most of which are based on assumptions that either the phenomena or physical properties of the things being modeled, complexity, or highly stochastic nature of the outcome is too nebulous to measure accurately and consistently. Although this is true in some cases, they can be successfully managed and supported by effective VV&A efforts as long as the sponsor recognizes the limitations and constraints imposed by the results. Some would call this "best effort" while others would systematically divide the problem space into parts that could be effectively and reasonably validated and those that could not. Evidence and test results would be collected for as much of the problem space as possible and subjected to more traditional forms of validation and analysis. Meanwhile, a team of SME and other experts would be convened to perform face validation and perhaps perform advanced statistical analysis on the more difficult remaining parts of the simulation or model.

The point is that the V&V agent has to know where the point of diminishing returns occurs and should try to stop just short of passing it. Because this estimating process has several very effective checks and balances built in, the sponsor can readily see what the VV&A effort is proposing to do for the assigned budget (in other words, every task is defined and scoped). If the sponsor wants to increase the V&V effort in specific task areas to help overcome these difficult validation issues, so be it. This can be accomplished in one of two ways:

1) Specific Validation tasks, LOE, and costs are added under the Validation Phase column in the Extraordinary Costs worksheet. This is the preferred method because it does not impact other VV&A tasks, still provides a reasonable plan and estimate, and allows all of the extraordinary validation activities to be tracked, managed, and accounted for separately.

2) If the extraordinary validation activities can be rolled into the pre-defined validation activities by maximizing the adjusted counts through tailoring, this is also acceptable. It does, however, impose more effort on behalf of the planner, who must enter data, examine the results, and tune the estimate until the correct answer is derived.

Both options above require that the planners prepare a separate estimate for this extraordinary validation support, which is another reason Option 1 above is preferred, since it is much easier to incorporate these figures into the estimate and V&V plan. Regardless of the option used, the VV&A planner is cautioned not to be carried away with an overly ambitious validation effort. If the developer built the M&S in the first place, but the validation looks very difficult, there may be some flaws in the assumptions concerning how to perform the testing and validation effort or what must be done to provide essential credibility. Full cooperation is required among all the participants to come up with a balanced, leveraged, and cost-effective validation plan and estimate. If the sponsor studies and accepts the V&V Plan with the levels of effort proposed and the V&V agent performs the specified work, yet there still remain un-reconciled validation objectives, it may be that the goals are unrealistic within the available resources. Use of outside experts, SME, and additional budget to extend the validation effort are all possible options. Option 1 above is also preferred when these shortcomings in the validation effort are discovered during testing. It is easier to add a separate task than to rework the entire V&V plan and estimates.

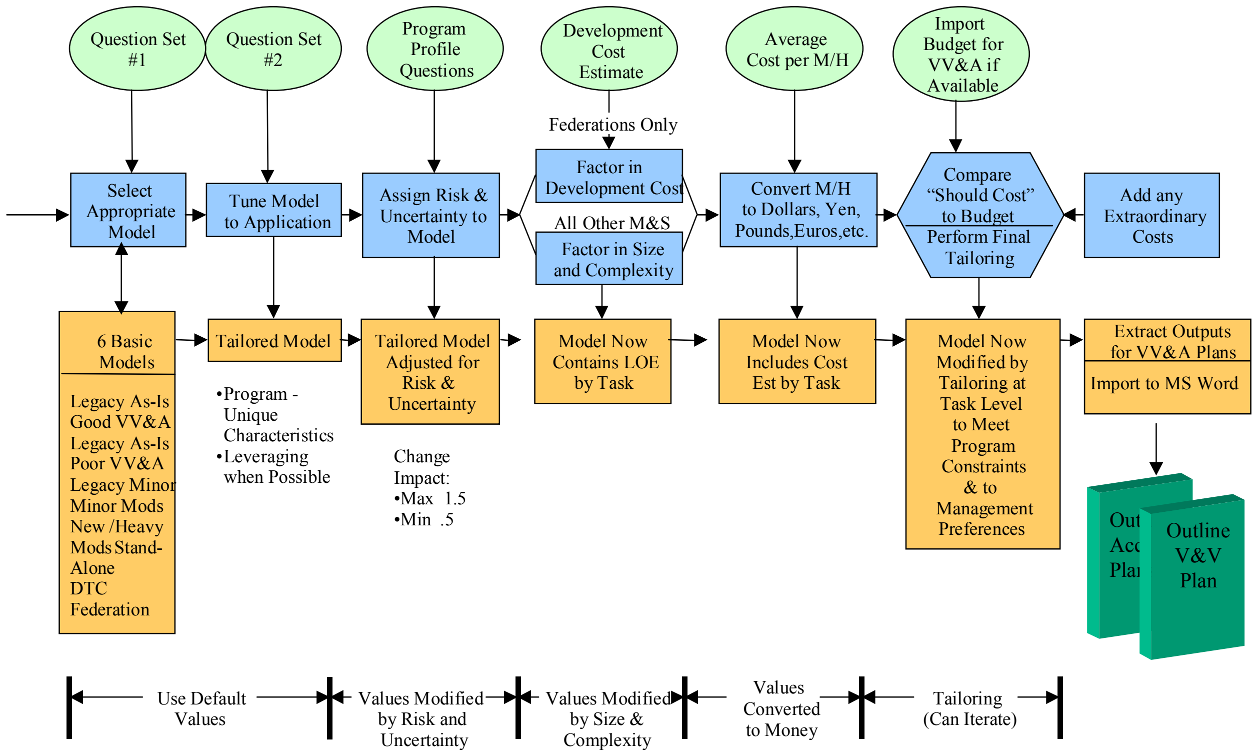
The fact remains that there will always be a few isolated examples of very difficult validation problems that require some serious planning, study, and collaboration among the developer, sponsor, user, and V&V agent to ever come up with a reasonable, workable solution. Along these lines, if the thing being modeled is so poorly understood that it cannot be validated against some referent, then the model may be a poor choice or an inadequate representation or implementation, and its use should be discouraged. When dealing with stochastic and chaotic systems, there are powerful mathematical tools that support the modeling itself and its validation; these should support both efforts. It also appears that modeling of human behavior falls into this category and can handled in much the same way as the more traditional M&S discussed thus far. Psychologists and Human Engineering experts may be brought in to review human and human surrogate behaviors.

A final thought involving difficult validation problems is that all the testing efforts conducted by all the participants combine to form the basis for validation. The V&V agent needs to be able to pull from all these sources those things that are most germane to the ultimate credibility and accreditation of the M&S product, regardless of its type and application. To this end, sometimes the best solution simply is to know where to stop. This decision should be reached mutually by all the participants in the program so that everyone agrees, and no one group feels burdened by too much responsibility for the success of the program. Remember that every time the M&S is used, more validation occurs.

Overview of the Cost Estimating Process and Tool

The following figure depicts the generalized cost estimating process, which is embedded in the CET. The steps below briefly describe the estimating process:

Step 1: Once the prerequisites are completed, there is a short set of questions that determines the appropriate VV&A model—legacy with or without modifications, new stand-alone, or federation. Once these are answered, the program takes one of five internal paths tuned to each type of model application. As additional models are added, the number of internal paths will increase accordingly.



Cost Estimating Tool Process Flow Diagram

Step 2: The second set of questions allows the CET to perform fine tuning of the VV&A application model selected in the previous step. In this step, consideration is made for the quality and completeness of the documentation, conceptual model, interface definitions, input data, leveraging, etc. that can affect the cost of the V&V effort. The more rework required by the V&V agent, the higher the cost.

Step 3: Next, the user of the CET is required to select his/her best estimate of the M&S program R&U factors based on a 15-question matrix. Values selected can influence the adjusted raw counts by as much as + or – 50%; however, those ranges are seldom experienced. Typical increases and decreases range between 5-15%. The output of the tool at this point is expressed in adjusted counts from the previous phase as modified by the risk and uncertainty calculations. Risk and uncertainty propagates down through the VV&A activities for each application model in a unique way. Everywhere that the R&U issue impacts a particular VV&A activity, the internal tool matrix gets a "hit." The user does not have to do this; it is already pre-determined. High impact or high priority R&U issues get more hits, making the weighting greater.

Steps 4 and 5: Now the development cost for federations or the size and complexity factors are introduced based on the languages being used in the software. The first resulting set of calculations is an estimated LOE for every task. The tool user enters the average cost per man-hour and the tool converts the LOE to cost in whatever money system is desired. The default is U.S. dollars. The total is the "should cost" labor figure and is available by activity, phase, and total program.

Step 6: Almost always, a VV&A effort will have to pay for extra cost items that are not included in the labor estimate. These are entered into the overall cost spreadsheet in two forms—phase specific and non-phase specific costs. It is here also that extraordinary costs pertaining to any feature of the VV&A effort should be added and tracked separately. It was mentioned earlier that if the M&S appears to be particularly difficult to validate, and the traditional validation activities in the VV&A application model do not seem adequate to support the effort, this is the place to add an independently planned and estimated adjunct validation effort. Having it here also provides a mechanism to track its progress and costs separately from the other more predictable in-line VV&A activities without distorting the plan. This is definitely the recommended way to handle difficult VV&A planning and estimating problems. It is not appropriate to load the normal VV&A cost estimate with these numbers unless the PM recommends this action, and is able to explain why the V&V costs are so disproportional to the average for this type of application. When such an extreme case is being considered, the effort should be priced separately and not merely buried in the overall VV&A cost without the full knowledge by all participants and disclaimers to the VV&A effort that explain the reason for the unusual cost elements. Otherwise, VV&A gets an undeserved reputation for fostering high cost efforts, and has no way to deny the allegations if such costs are anonymously folded into the bottom-line cost. Every effort should be made to protect VV&A from these types of accusations. After extra costs are delineated, the total cost of the VV&A program is calculated.

Step 7: The final tailoring occurs when and if the sponsor has a number in mind for VV&A that disagrees with the "should cost" figure derived from the CET. The tool user can now adjust the VV&A activity matrix to reach the budget number, if directed to do so. The R&U values can be changed, the leveraging can be changed, tasks can be raised or lowed in scope and LOE, etc. to converge the numbers. However, if the VV&A tool user has to go more than 15% to reach a match by *reducing* the VV&A effort, the true risk to the program will go up at least that much. In such a case, the tool developer urges the tool user to recheck all the values, call the designer of the tool if necessary for confirmation and discussion, and then take the results to the sponsor for final in-depth discussions and resolution. Having any extraordinary cost as a separately accountable item is a very effective way to negotiate the best value, least-risk VV&A effort regardless of the budget.

Saving the Best Part for Last

The CET is available free of charge to anyone who wants it by contacting the developer, Robert O. Lewis.

Case Histories

A number of programs have considerable experience dealing with the costs and schedule impacts of M&S V&V activities. Some of those programs have agreed to tell their stories in this paper in order to shed light on their actual experiences in applying VV&A processes and the resulting resource expenditures. Almost all of these programs are from the Department of Defense, because of the nature of the authors' experience and expertise.

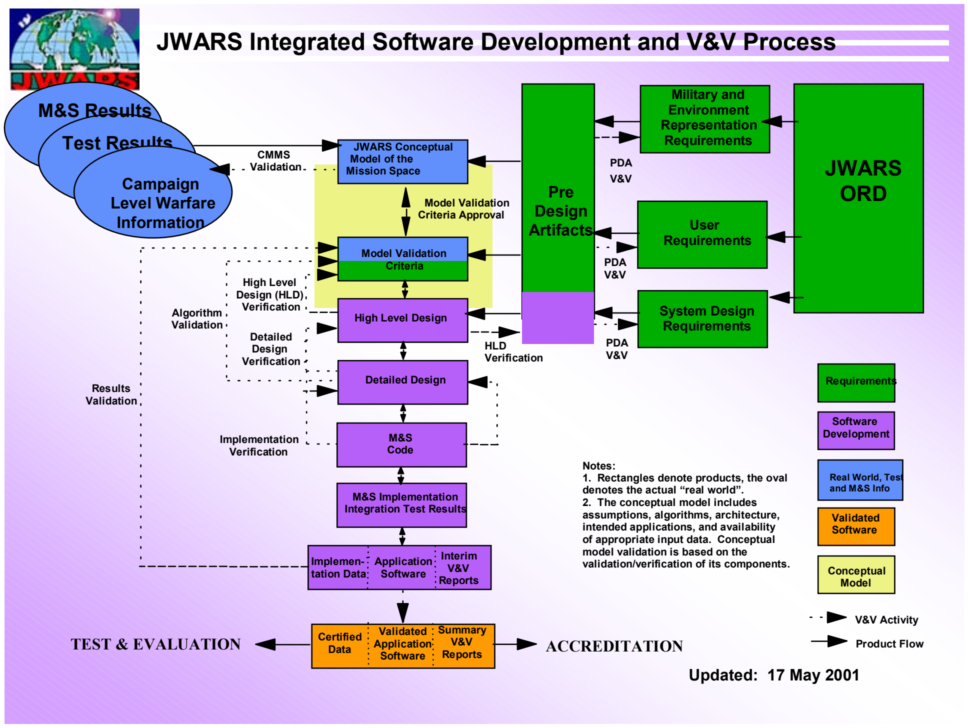
A Case Study – the Joint Warfare System (JWARS)

The JWARS V&V effort began in late September 1997 with the award of a contract to JWARS V&V, a joint venture of Innovative Management Concepts, Inc., (IMC), in Sterling, Virginia, and BMH Associates, Inc., (BMH) in Norfolk, Virginia. Both IMC and BMH had extensive experience in military simulation development and V&V. As the result of the contract award the BMH-IMC Team was selected as the JWARS V&V Agent.

The JWARS Office is located in the Office of the Director, Program Analysis & Evaluation (OD PA&E), in OSD. With their approval, the JWARS V&V Team formed an Integrated Product Team (IPT) of the interested stakeholders in JWARS in order to develop the JWARS V&V process, direct the writing and coordination of the JWARS V&V Plan, and manage the execution of the plan.

The JWARS Office and the contract statement of work provided the following guidance in the development of the JWARS V&V process:

- The process must follow the existing JWARS software development process (SDP) and the V&V Agents must not ask for new or additional development artifacts.
- V&V activities should be coordinated with the JWARS Iteration 1 through 9 schedule and there should be a report for each Iteration
- The process must be based on and compatible with the then existing *DOD VV&A Recommended Practices Guide.*
- The V&V Agent must receive approval for the process and plan from the Joint Analytic Model Improvement Program (JAMIP) Steering Committee and report regularly to the committee.
- **The five-year cost of the effort could not exceed $2.5 million**.



Original JWARS Simulation Development Process and V&V Process

In October 1997 the JWARS Office created the JWARS V&V Oversight Group made up of representatives from the JWARS Office, the Developers, DMSO's VV&A Technical Director, the Washington area based Service analysis agencies (Army Concepts Analysis Agency -CAA, Air Force Studies and Analysis Agency-AFSAA, Navy N-81, and Marine Corps Combat Development Command-MCCDC), the OD PA&E Studies and Analysis Center (SAC), the Joint Staff J-8 and J-4, Federally Funded Research and Development Centers (FFRDC - RAND, IDA and Mitre), the Joint Data Support organization, and the V&V Agent.

The Oversight Group met frequently during the first month at the JWARS Office and explored the basic V&V tasks that needed to be accomplished and developed a rudimentary set of validation and verification tasks. The group identified the more than 1500 members of the JWARS User Sub Groups as the validation authority for JWARS warfare functionality and the V&V Agent as the Verification Agent. The V&V tasks that were identified included:

- Conceptual Model of the Mission Space (CMMS) validation
- Conceptual Model Validation
- Algorithm Validation
- Design Verification (both High Level and Detailed Design)
- Code Verification
- Implementation Verification
- Results Validation

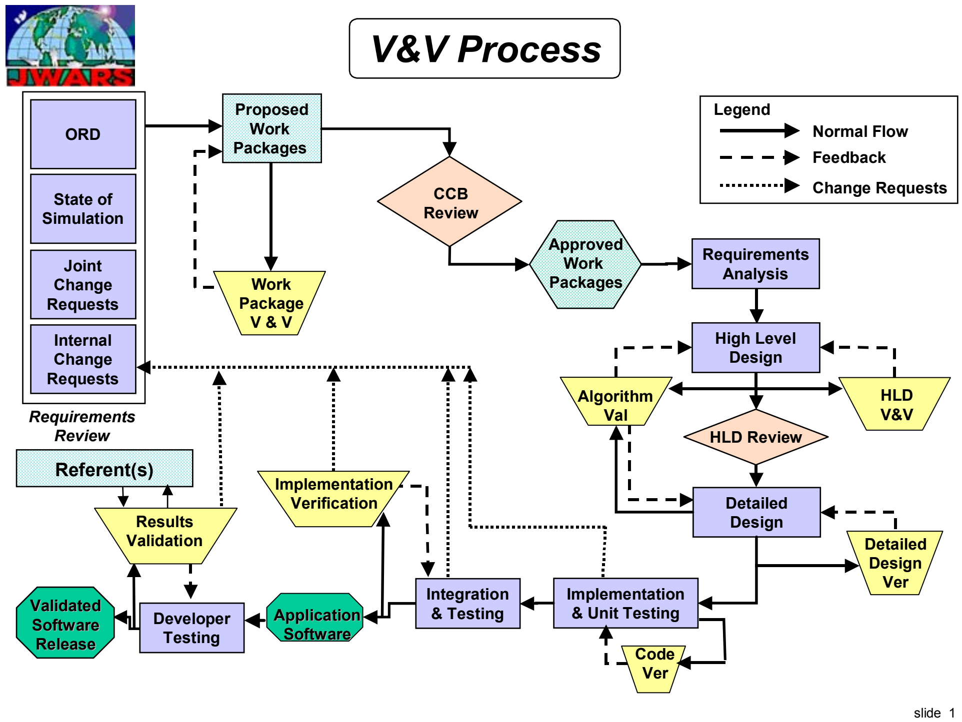
The V&V Agent then created a matrix of the labor required to conduct each task, for each of the scheduled iterations (each of the iterations was scheduled for about six months). The result of this analysis showed that there were insufficient funds to do each of the planned tasks for all iterations. The Oversight Group determined that the Code Verification and Implementation Verification tasks were duplicative of the Developer Testing and the planned Operational Testing and could be reduced or eliminated.

The V&V process was completed and presented to the JAMIP Steering Committee. The committee accepted the process and directed that all V&V activities be coordinated with the operational testing and evaluation (OT&E) activities to reduce the impact on the Developers and to maximize the synergy between the two efforts. They also directed that the V&V Agent utilize, to the maximum extent possible, all JWARS Developer Testing and Evaluation (DT&E) processes and products.

The Oversight Group was renamed the JWARS V&V – T&E Working Integrated Product Team (WIPT) in 1998 and both the DT&E and OT&E processes and plans were added to the agenda as regular items.

During 1998 the WIPT met twice a month to review the V&V process and the V&V plan that was written by the V&V Agent. The WIPT approved the plan for coordination and the JWARS Office sent it out for comment and approval. During this process several hundred comments and suggestions were received. These were compiled by the V&V Agent and adjudicated by the WIPT. At the conclusion of this process the JWARS V&V Plan, Version 3.0, was approved by the JAMIP Steering Committee.

In 2000 the JWARS Office began to revise the simulation development process away from the then existing Joint Application Designs (JAD) as the primary pre-design artifact, and to the use of work packages as the primary pre-design artifact. While the JAD were based on relatively high level processes and took six months or more to design and program, the work packages were much smaller and could be designed and programmed in approximately one month. The revised process and associated V&V process is shown in the Figure below.



Revised JWARS Simulation Development Process and Associated V&V Process

Funding and Schedules

JWARS V&V was funded under a Cost-Plus-Fixed-Fee contract issued by the Defense Supply Service – Washington, a Contractor Officer's Representative in OD PA&E, and the Technical Monitor in the JWARS Office.

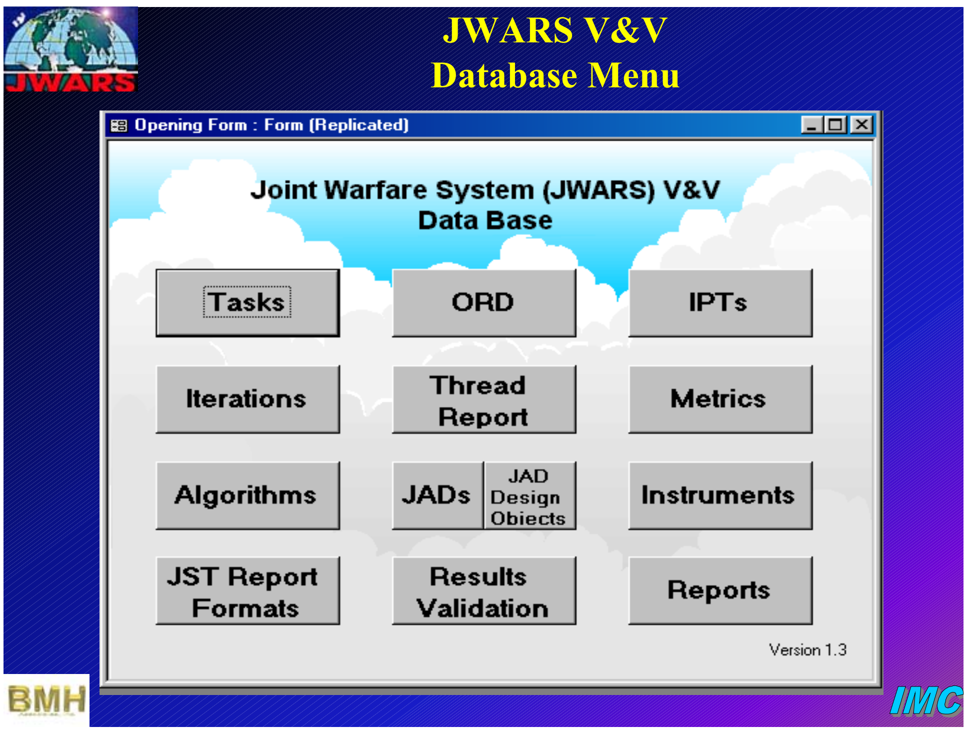
Delivery Order Schedule and Cost per Delivery Order for the Entire Contract

Delivery Order #	Start	End	Dollar Value
1	30 September 97	9 August 1998	$279,254
2	10 August 1998	9 August 1999	$489,674
3	10 August 1999	9 May 2000	$387,348
4	10 May 2000	11 March 2001	$494,566
5	12 March 2001	12 November 2001	$363,099
6	13 November 2001	12 November 2002	$557,050
Total			$2,570,991.00

The full-time equivalent (FTE) personnel level began at two people for delivery order one and increased to 3.7 for subsequent delivery orders. The V&V report schedule is shown below:

Release .5 (alpha) V&V Report – September 1999
Release 1.1 V&V Report – September 2000
Release 1.2 V&V Report – September 2001
Release 1.3 V&V Report – April 2002
Release 1.4 V&V Report – estimated November 2002.

In addition to providing the required Release V&V Reports, the V&V Agent also developed the JWARS V&V Database. The JWARS V&V Database provided an excellent tool for maintaining all requirements, archived development artifacts, and V&V activities. The "JST" shown in the Database menu below was the JWARS Study Team, a group of early JWARS users formed to wring out JWARS and suggest fixes and improvements; members of the JST were from Army CAA, Navy N-81, Joint Staff J-8 and J-4, OSD Studies and Analysis Center (OSD/SAC), Air Force Studies and Analysis Agency, and the Missile Defense Agency.



JWARS V&V Data Base Menu

The status of the JWARS Release 1.4 activities is shown in the table below, as of June 2002.

V&V Activity	Artifact Status	Start	End
Work Package Validation	Provided by JWARS Office NLT 30 June 02	1 July 02	1 Aug 02
Design Verification	Provided by JWARS Office NLT 30 June 02	1 July 02	31 August 02
Results Validation	Release 1.4 available 30 April 02		
- Base Case Analysis		1 May 02	30 June 02
- Sensitivity Analysis		1 June 02	31 July 02
- TACWAR Comparison	With SAC concurrence [1]		
- Objective SME Reviews	V&V Agent must build mini scenarios	1 May 02	15 October 02
Write 1.4 Report (Draft)		15 Sept 02	15 Oct 02
Write 1.4 Report (Final)		15 Oct 02	12 Nov 02

Cost and Schedule Impact of the JWARS V&V effort on JWARS Development

The cost of the JWARS V&V effort will be $2,570,991 through the first five years of the effort, specifically for the V&V Agent's labor, travel, hardware, and software. However, there have been other costs associated with the V&V effort. One of these is the cost of providing technical direction to the V&V Agent by JWARS Office government personnel. Although no formal measurement has been done of how much time the JWARS Office spends working with the V&V Agent, it is estimated that the government person providing that direction (a Navy Commander) has spent at least 10% of his time working V&V related actions.

A second cost is the time required by members of the JWARS Office and the other members of the JWARS WIPT in preparing for and attending the WIPT meetings. Several thousand hours of government and contractor time have been provided to support this effort. However, the role of the WIPT expanded in time to include test and evaluation, external support, developmental testing, and operational testing so all of these thousands of hours cannot be directly attributed to the V&V effort.

Another cost to the JWARS effort has been the changes made in the simulation development process and or artifacts made at the recommendation of the V&V Agent. Among these recommendations, all of which have been accepted by the JWARS Office as improving the JWARS development effort and providing development risk reduction, are:

- Standardizing the pre-design artifacts (JAD and Work Packages) format and content
- Providing improved algorithm documentation

- Improving design processes and products
- Developing an expanded Analyst and User manual as a JWARS Conceptual Model
- Correcting deficiencies identified by the V&V Agent during V&V testing

Each of these changes required additional work by the JWARS development contractors. However, they were requested to both improve the JWARS simulation development process (bringing it more into line with accepted software and simulation development practices) and provide the artifacts needed to conduct the V&V effort. The V&V Agents believe that they did not go outside of their direction (which was not to ask for artifacts strictly for the purpose of V&V) as they were convinced that the changes they requested first made the JWARS product better and second provided the necessary items for the conduct of the V&V effort.

A Case Study – Using A Managed Investment Strategy for VV&A of MDA's Ground-Based Radar Prototype (GBR-P) HWIL Testbed

The Managed Investment Strategy and associated methodologies for M&S VV&A planning and execution have been successfully used by AEgis Technologies for several major simulation programs within the acquisition and test domains. This strategy has been, or is being used for a wide variety of constructive and virtual simulations including:

1) Boeing Company's Prime Consolidated Integration Laboratory (PCIL) for the Ground-based Midcourse Defense Element;
2) Ground Based Mid-Course Defense Element's Integrated System Test Capability (ISTC);
3) US Army Aviation and Missile Command's Patriot Advanced Capability III (PAC-3) Millimeter Wave Simulation;
4) Missile Defense Agency's Missile Defense System Exerciser (MDSE); and
5) MDA's Ground-Based Radar - Prototype (GBR-P) HWIL Testbed.

The VV&A effort for this last simulation program [GBR-P HWIL Testbed] has been completed and will serve as a use case to illustrate the application of a "Managed Investment" strategy for M&S VV&A in greater detail and to demonstrate its utility, value; and extensibility in developing similar M&S assessment and VV&A programs.

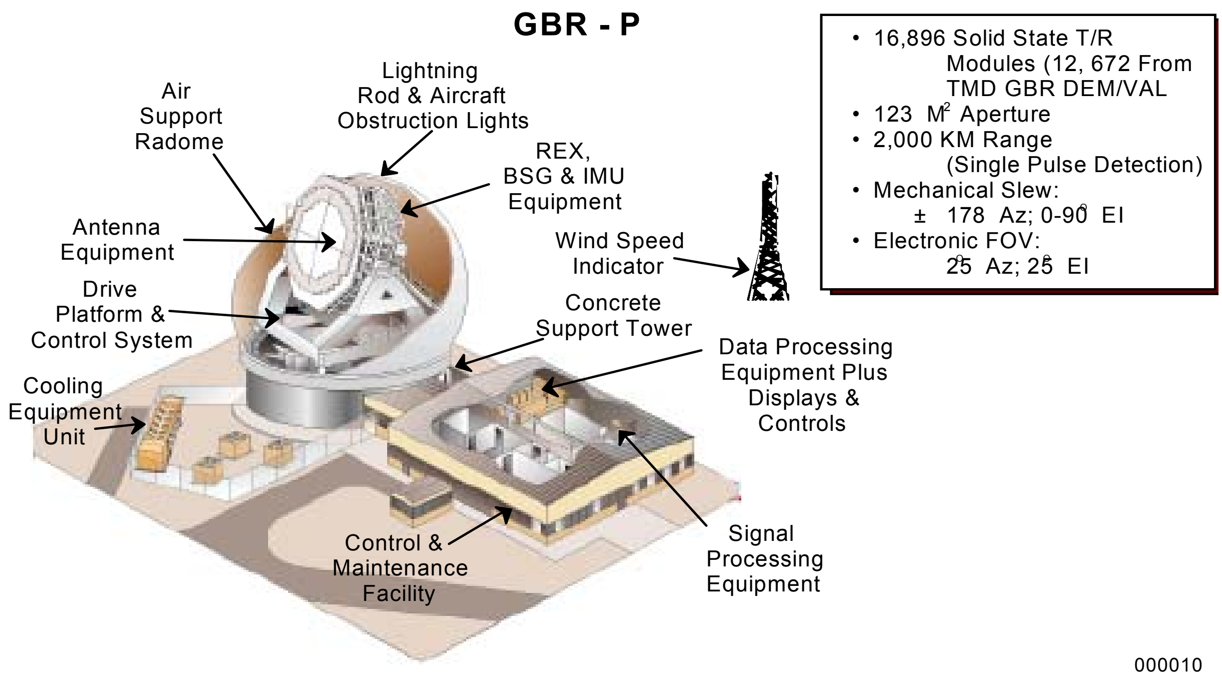
Background On GBR-P and the GBR-P HWIL Simulation Testbed

The GBR-P is an Anti-Ballistic Missile (ABM), treaty-compliant, test radar built to resolve and demonstrate critical technology issues for the X-Band Radar (XBR) element for the Ground-Based Midcourse Defense (GMD) Segment. The GBR-P Radar System is located at US Army Kwajalein Atoll (USAKA), serves as a key test resource for resolving critical technology issues for the XBR, and supports integrated GMD weapons system testing.

The primary objective of the GBR-P was to perform surveillance, acquisition, tracking, discrimination, interceptor support, and kill assessment in the exo-atmospheric threat flight regime. The GBR-P performs coordinated operations and interfaces with external installations via the Kwajalein Mission Control Center (KMCC) and the BM/C3. These interfaces permit the GBR-P to accept handover information from external sensors. The GBR-P can be operated based on external tasking, or autonomously under the control of the system operator. The layout of the GBR-P is illustrated below.

The GBR Project Office, concurrent with the fabrication of the GBR-P radar, developed the GBR HWIL simulation testbed. Additional technical and programmatic direction was provided by the National Missile Defense Joint Program Office (NMD JPO) [renamed the GMD JPO] and the Ballistic Missile Defense Organization (BMDO) [now the MDA].

Developing the GBR HWIL simulation testbed entailed active coordination with US Army, Navy, Air Force, BMDO and DoD agencies as well as with collateral, government and contractor-developed test beds and simulations.



Overview of the Ground Based Radar – Prototype at Kwajalein Atoll, US Marshall Islands

With delivery of an initial operational capability in July 1996, the GBR HWIL simulation testbed supported developmental and operational testing, as well as material developers, combat developers, and operational commanders. The GBR HWIL simulation testbed provided the capability to define, execute, and collect HWIL and software-in-the-loop (SWIL) simulation experiments over a wide range of GBR-P system design, test and evaluation, and operational areas of interest. The scope of systems represented in the GBR HWIL simulation testbed included ballistic missile threats; air-breathing threats; satellite-, ground and air-based sensors; electronic countermeasures (ECM); surrogate battle management and command, control and communications (BMC3); pertinent weapons effects environments; and terrain and atmospheric phenomena.

The GBR HWIL simulation testbed consisted of a "simulation framework", a set of generic "common models, as well as a variety of GBR-P system-specific and GBR-P component representations of corollary, real-world GBR-P radar components. The common models provided the "methods" for representing generic object classes from which representations of real-world systems were composed. These representations were comprised of common or legacy models, their associated "characteristics" (parametric) and "instance" (initial condition) data, and decision processor "rule sets". The V&V activities had to confirm both that the generic model methods were correct, and that the components and systems composed from them adequately represent the "real-world" prototype system.

The first issue regarding accreditation of the GBR HWIL simulation testbed was its scope of application. DOD and Service direction indicated that its accreditation was contingent on application scope (intended use), that the application domain must be specified explicitly, and that it could be incrementally accredited for a progressively wider scope of application. In addition, the Department of the Army guidance recognized that accreditation could be conferred for either a particular study, or for a "class of applications". While study managers were to be responsible for conducting study-specific accreditation, the accreditation for classes-of-application were to be managed by the simulation developer / sponsor, which in this instance was the GBR-P Project Office. The classes of application for which the GBR-P HWIL simulation testbed was expected to eventually apply include those indicated in the Figure.

Consequently, the GBR-P HWIL V&V activities were selected to:

1) Accredit the GBR HWIL initially by the GBR Project Manager as a test resource for SWIL tests, coupled with ground and flight test pre-mission and post mission support functions; and
2) Establish the simulation testbed as a functional and performance representation of the GBR-P radar at USAKA through execution of additional V&V activities focused on this application.

- **ANALYSIS**
- **TEST AND EVALUATION**
- **EDUCATION AND TRAINING**
- **PRODUCTION AND LOGISTICS**
- **RESEARCH AND DEVELOPMENT**

Possible GBR-P HWIL "Class of Applications"

Within resource and schedule constraints, additional V&V activities were subsequently undertaken to support accreditation of the GBR HWIL by:

1) Other Accreditation Agencies interested in using the GBR HWIL to support their independent assessments of the GBR system; and
2) BMDO in support of an NMD contingency deployment decision in calendar year 2000.

Another consideration in specifying the M&S accreditation data information requirements was to establish the level and span of authority of the Accreditation Agent. In the case of the NMD GBR HWIL, a joint-service accreditation at the level of the Director, System Test and Evaluation, NMD Joint Project Office was considered appropriate, notwithstanding the additional difficulty in coordinating a joint-service decision. The management and administrative mechanism for a joint-service accreditation of the NMD GBR HWIL Test Bed for a specified class-of-applications was in place via the NMD T&E Resources VV&A Advisory Panel[18] and the NMD T&E Program Integrated Product Team (PIPT) that existed at that time in the program.

It was envisioned that the GBR Project Office, with AEgis' V&V support, would execute the GBR HWIL V&V plan, generating records of original entry and preparing reports and

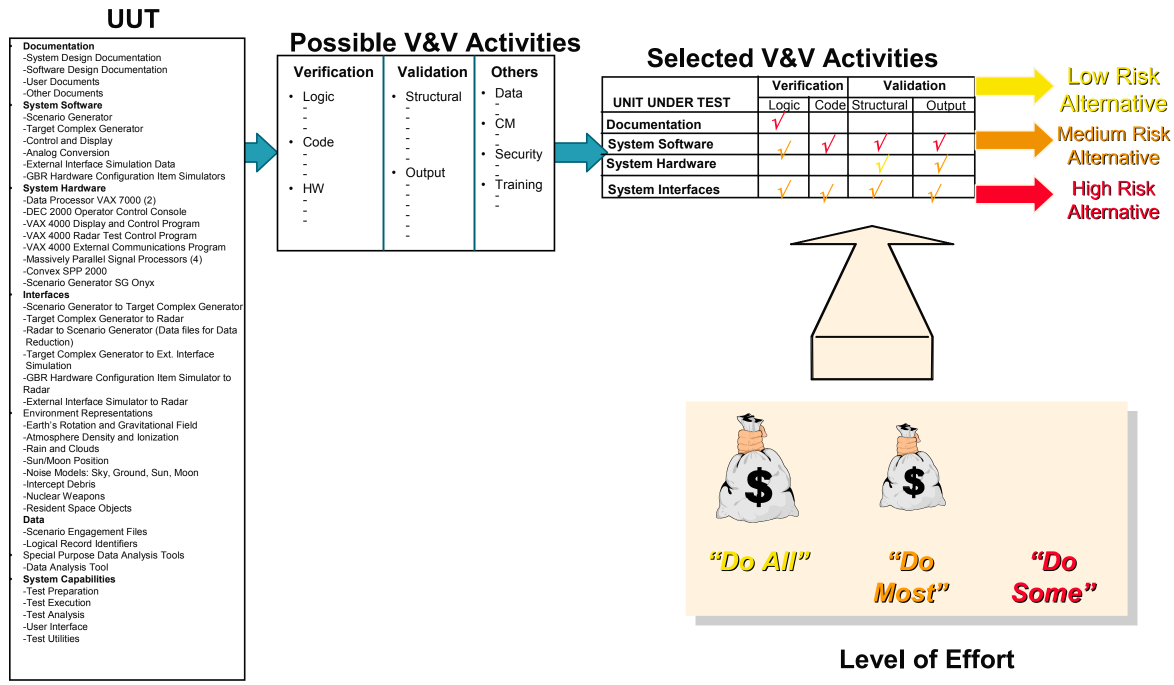
[18]. The NMD T&E Resources VV&A Advisory Panel was a high level group chartered by the Director, System Test and Evaluation, NMD JPO with membership including senior representatives of government agencies and activities that were most likely to be potential Accreditation Authorities.

necessary abstracts to support accreditation decisions. These materials were made available to the NMD T&E Resources VV&A Advisory Panel. It was anticipated that the NMD T&E Resources VV&A Advisory Panel would periodically review and provide comments and suggestions throughout the planning and execution of the NMD GBR HWIL Test Bed VV&A program. The NMD T&E Resources VV&A Advisory Panel, under the leadership of the Director, System Test and Evaluation (T&E), NMD Joint Project Office, made recommendations regarding potential BMDO and broader DoD accreditation decisions, and provided notification to the NMD T&E PIPT.

For the GBR HWIL Testbed there existed models of a wide variety of entities, which needed to be verified and validated in support of the accreditation determination. This circumstance required a disciplined development and V&V process and adequate supporting documentation. The technical approach to addressing these concerns included:

1) Careful identification of the GBR HWIL UUT;
2) Diligent distribution of V&V effort across the GBR HWIL UUT; and
3) Explicit qualification of results.

Potential UUT that were similar in nature were grouped for the efficient execution of selected V&V activities. Some V&V activities were considered not feasible or practical, and therefore not recommended for execution. The remaining, feasible V&V activities were then individually evaluated to establish an estimated LOE and cost of execution. The LOE and associated costs for certain V&V activities were then reduced, or in some instances eliminated altogether as the high cost-low risk program was adjusted to a medium cost-medium risk program, and then further reduced in scope for a low cost-high risk program . Each of these alternative programs is summarized below.



Alternative V&V programs were derived for the GBR HWIL Testbed using the Managed Investment strategy discussed earlier in this paper. Tables summarizing the suite of V&V activities associated with the GBR HWIL and their associated costs for execution were generated using COTS software to "spreadsheet" the V&V effort over a three-year period of performance. Illustrated below are three perspective views of V&V activities that could be conducted during execution in which the risks of non-accreditation were ranked as low, medium or high, but the corresponding costs of executing the V&V activities were ranked as high, medium or low. Each of these alternatives is described in greater detail below.

GBR High Cost-Low Risk Alternative V&V Program Cost Summary.

	FY97		FY98		FY99	
	MW	Cost ($)	MW	Cost ($)	MW	Cost ($)
Documentation	12	24	17	34	10	20
System Software	76	152	176	352	60	120
System Hardware	18	36	35	70	29	58
Interfaces	8	16	50	100	40	80
Environmental Models and Data	106	212	220	440	110	220
Sp Purpose Tools and System Capabilities	29	58	63	126	22	44
TOTAL EFFORT BY YEAR	**249**	**498**	**561**	**1122**	**271**	**542**

Note: Current Year $ in Thousands

	High Cost-Low Risk Alternative
TOTAL FUNDING REQUIRED	**$ 2,162 K**

GBR Medium Cost-Medium Risk Alternative V&V Program Cost Summary.

	FY97		FY98		FY((	
	MW	Cost ($)	MW	Cost ($)	MW	Cost ($)
Documentation	7	14	9	18	6	12
System Software	47	94	77	154	35	70
System Hardware	7	14	12	24	13	26
Interfaces	7	14	32	64	22	44
Environmental Models and Data	62	124	117	234	63	126
Sp Purpose Tools and System Capabilities	21	42	41	82	12	24
TOTAL EFFORT BY YEAR	**151**	**302**	**288**	**576**	**151**	**302**

Note: Current Year $ in Thousands

	Medium Cost-Medium Risk Alternative
TOTAL FUNDING REQUIRED	**$ 1,180 K**

GBR Low Cost-High Risk Alternative V&V Program Cost Summary.

	FY97		FY98		FY99	
	MW	Cost ($)	MW	Cost ($)	MW	Cost ($)
Documentation	7	14	5	10	3	6
System Software	33	66	48	96	21	42
System Hardware	6	12	13	26	3	6
Interfaces	4	8	22	44	4	8
Environmental Models and Data	26	52	88	176	28	56
Sp Purpose Tools and System Capabilities	6	12	24	48	8	16
TOTAL EFFORT BY YEAR	**82**	**164**	**200**	**400**	**67**	**134**

Note: Current Year $ in Thousands

	Low Cost-High Risk Alternative
TOTAL FUNDING REQUIRED	**$ 698 K**

Alternative VV&A Programs for the GBR HWIL Simulation

Lowest Risk (Highest Cost) V&V Program

This served as the baseline case and was the departure point for other V&V investment program alternatives. This proposed V&V program was a complete, broad, and quite thorough program of assessment activities to lower risks associated with obtaining a favorable accreditation decision. All of the V&V activities selected for execution were estimated with sufficient LOE necessary for an in-depth assessment. The resultant cost to execute the proposed Lowest Risk - Highest Cost V&V Plan was approximately $2.162 million over three years. This could have been viewed as excessive in scope, and not “realistic”, in today’s economically constrained defense budget. The risk to the GBR Project Office in not executing a comprehensive V&V program for the GBR HWIL Testbed was the potential of it not being accredited, and unable to support a production or contingency deployment decision; e.g., avoidance of a Type I error[19]. Although it was recognized that the V&V effort needed to be systematic, it could not, and would not be exhaustive. Therefore careful allocation of resources onto high-return-on-investment V&V activities was critical. A particularly significant issue was the iterative balancing of the program scope against available resources to ensure that the investment in V&V activity provided the best possible return on investment.

Medium Risk (Medium Cost) V&V Program

This alternative investment provided a perspective view of the feasible V&V activities that could be conducted during execution for a program in which the risk of non-accreditation was higher than the baseline, but the associated costs of executing the V&V activities were reduced. The risk of non-accreditation and associated costs for the V&V activities of this alternative were between the low-risk, high cost alternative and the high risk, low cost alternative presented to the GBR-P PM for consideration. The V&V activities were reduced in scope when compared to the lowest risk, highest cost program.

Reduction in the level of effort for this program alternative was based upon trade-off analyses of the consequences of executing the proposed V&V activity and the value of that information in supporting an accreditation determination. Cost as an independent variable was a key consideration in this trade-off analysis. Additional activity selection factors and level-of-effort loading criteria included:

- Accreditation agency information requirements;
- Historical perspective and lessons learned from other M&S VV&A programs;
- Incorporation of activities having little or no cost impact because they were already embedded and funded as part of the system/software design process, like software design reviews; and
- Application of engineering judgment.

The resultant cost to execute the proposed medium-risk, medium-cost V&V Plan was $1.18 million over three fiscal years.

High Risk (Low Cost) V&V Program

This suite of activities represented a cursory V&V program that may lead to the successful accreditation of the GBR HWIL Testbed, but with substantial risks. The risk of non-accreditation for the GBR HWIL Testbed was much higher with this minimum set of activities, but the associated costs for this suite of V&V activities was substantially lower than the higher cost alternatives presented to the GBR PM.

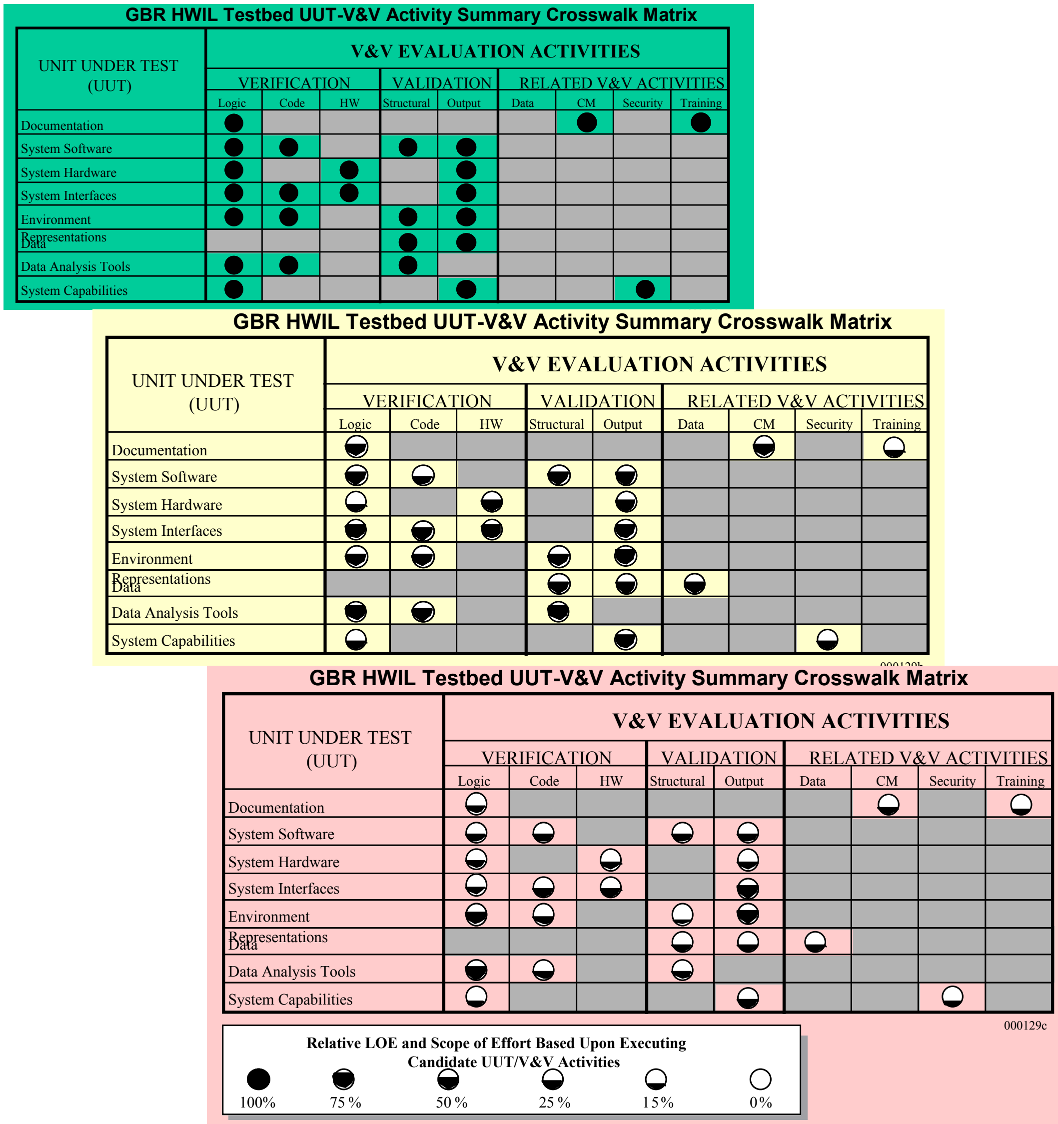
[19] A Type I Error occurs when simulation results are rejected although in fact they are sufficiently credible. Committing this type of error unnecessarily increases the cost of M&S development. The probability of committing this type of error is referred to as Model Builder's Risk. To avoid a Type I Error, application requirements must be very clear, and the simulation results must be carefully considered against the requirements

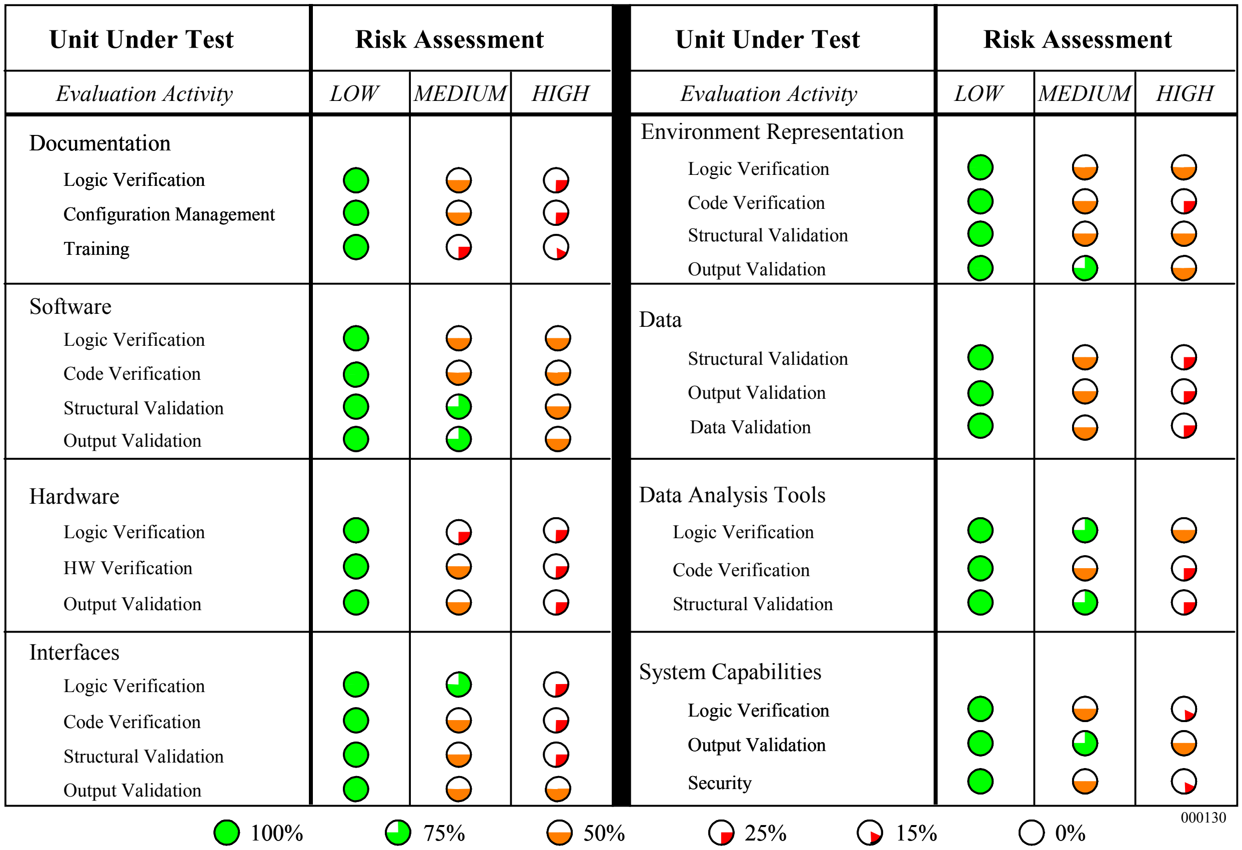
This alternative was the most resource constrained of the three alternative V&V programs and was considered marginal in obtaining the information necessary to support an accreditation determination for the GBR-P HWIL Testbed. The costs estimated for this high risk, lowest cost V&V program was only $698K over three fiscal years.

Evaluation of Investment Alternatives

A wide variety of entities in the GBR-P HWIL Testbed needed to be verified and validated. This circumstance required a disciplined V&V planning and investment process to establish the nature of the program that could be supported within the available resources. A series of tables, illustrated below, depicted the associated investment alternatives provided to the GBR PM. The V&V activities selected for execution for each of the alternative V&V programs were individually evaluated to establish LOE and costs of execution.

The LOE and associated costs for certain activities were reduced or in some instances, activities were eliminated altogether as the lowest risk, highest cost program was adjusted to a medium risk, medium cost program, and then further reduced in scope for the highest risk, lowest cost program. These alternative programs of V&V activities could then be compared in terms of scope (evaluation activity), depth and breadth (investment), and associated risk of having insufficient information to support the accreditation decision. This comparison of alternatives is illustrated below.


GBR HWIL Testbed UUT-V&V Activity Summary Crosswalk Matrix
UNIT UNDER TEST (UUT)
V&V EVALUATION ACTIVITIES
VERIFICATION
VALIDATION
RELATED V&V ACTIVITIES
Logic
Code
HW
Structural
Output
Data
CM
Security
Training
Documentation
System Software
System Hardware
System Interfaces
Environment
Data Representations
Data Analysis Tools
System Capabilities
000129c
Relative LOE and Scope of Effort Based Upon Executing Candidate UUT/V&V Activities
100%
75 %
50 %
25 %
15%
0%



A Comparison of VV&A Program Alternatives for GBR-P HWIL Testbed

Subsequent trade-off analyses considered the costs of executing each proposed V&V activity and the value of that information in supporting an accreditation determination were conducted. Cost as an independent variable was a key consideration.

Observations and Conclusions

The potential costs and risks associated with the three V&V LOE for the GBR-P HWIL testbed during the FY97 to FY99 period have been clearly illustrated above. This program's experience and the associated methodology demonstrate the value of a managed investment strategy in the development and subsequent tailoring of a VV&A enterprise.

It is clear that the timeline of the GBR HWIL Testbed development effort and the resources available to the conducting agencies constrained what V&V activities were undertaken, what V&V products were generated, and the degree to which all the V&V needs of this program were met. Consequently, the GBR PM needed to execute a plan of V&V activity which was responsive to the needs of the BMDO accreditation authority and other potential accreditation agencies, and was practical to implement.

Funding established the available LOE for the V&V program and directly impacted the selection of V&V activities and their application to GBR HWIL units-under-test.

A Case Study: AIM-9X

The AIM-9X Sidewinder air-to-air missile development program is a "poster child" for the effective use of M&S in a Department of Defense system acquisition program. In fact, the AIM-9X program won an award in FY01 from DMSO for their use of M&S as part of development, and in support of test and evaluation. Two primary M&S used by the program are the Integrated Flight Simulation (IFS), a six degree-of-freedom missile flight simulation, and the Joint Services Endgame Model (JSEM), which models the performance of the missile fuze and warhead (and calculates an engagement probability of kill against air targets).

The AIM-9X program has been very diligent in its pursuit of credibility information to support the use of M&S in system development and testing. Both the IFS and JSEM were accredited by the Program Manager late in FY01 for use in support of system specification development. The Program Manager and Commander, Operational Test and Evaluation Force (COMOPTEVFOR) intend to accredit IFS and JSEM for use in demonstrating specification compliance, live fire test and evaluation (LFT&E), and OT&E. The approach the program has taken in each of these cases is to convene an accreditation review panel at several stages throughout the VV&A process; this has been done to gain concurrence on the overall approach to VV&A, to review V&V data as they come in, and to develop an accreditation recommendation to the program manager with all the supporting V&V information required to support a final decision. Separate validation review panels also have been convened specifically to investigate model validation opportunities associated with test firings of the missile.

The Joint Accreditation Support Activity, which acts as the M&S accreditation agent has assisted the AIM-9X program in planning for VV&A activities. The missile developer (Raytheon Missile Systems Company in Tucson, AZ) has primary responsibility for V&V of the IFS, while the Naval Air Warfare Center, Weapons Division (NAWCWD) at China Lake, CA has responsibility for V&V of the JSEM ModX version which was developed at NAWCWD specifically to support AIM-9X fuze and warhead analyses.

VV&A Cost Experience

V&V of the IFS was distributed over more than one contractor; Raytheon subcontracted part of the V&V documentation effort to Computer Sciences Corporation (CSC - who also was on contract to JASA for other tasking). The total V&V costs for IFS were spread over several contract iterations, and the direct deliverables attributable to the V&V tasking were not always well defined. V&V of JSEM was primarily accomplished by the NAWCWD developer, but some of the tasking was done by JASA and its support contractors (SURVICE Engineering Company and CSC). Thus the costs of VV&A, being distributed over several organizations and contract vehicles over several fiscal years, are difficult to reconstruct.

AIM-9X has been able to make use of live fire missile shots since early in the program to provide validation data for the IFS. Each of those shots has been examined in detail by a validation review panel, comparing the actual test results with predictions made using the simulation, and the results are factored back into the IFS as appropriate. Similar, but fewer test data opportunities have been available to JSEM via other testing approaches and facilities. Thus the validation effort for both M&S has been extensive. However, the costs of collecting the validation data have not been charged against the M&S VV&A effort (since the tests were conducted for other reasons), so it is difficult to say with any certainty what were the resource requirements associated with model validation as part of the test series, other than the costs of the validation reviews themselves.

It is difficult to list the total number of validation reviews except that there was one for each missile fired, whether it was a guided shot or a Separation/Control Test Vehicle (SCTV - specifically used to demonstrate that the IFS was working properly). Many reviews were combined, and there were a few reviews that evaluated the models of subsystems (e.g., the Imaging Subsystem) that used laboratory data as well as missile test flight data. The best estimate is that there were a total of about 20-30 validation reviews during FY01 and FY02. The only known cost data available for those reviews was from the CSC contract, which charged approximately ½ work-years for participation in (and facilitation of) those reviews. However, there were many others who participated in the reviews (as many as a dozen people at each review – representatives of the missile developer, the simulation developer, the program office, the test range, etc.), making the total resources required for those reviews larger then the known amount charged by the CSC person who coordinated them. Also not included are the time and effort on the part of the presenters to reduce the telemetry data, do the post-flight simulation runs, construct the overlays, annotate the overlays, and prepare the presentation material for the reviews.

The support provided by the JASA Accreditation Agent was approximately 1-2 work-years per year during the period FY99 through FY02. Tasking that this effort supported included ensuring that detailed V&V documentation requirements to support accreditation were identified and defined. JASA provided expert inputs to documentation requirements and templates in coordination with the model developers. They provided technical reviews of interim and final V&V and accreditation support products (documentation). The most visible function that the accreditation agent provided was support of M&S and Accreditation Reviews. This included recording minutes; developing and maintaining summary reports of ongoing V&V and accreditation support activities and status; evaluating the technical status of VV&A activities; keeping the Program Manger apprised of emerging or potential risks to accreditation; and providing recommendations for risk mitigation as appropriate to all the players. Total cost of this effort over four years (FY99 thru FY02) was on the order of $1.5M.

The cost of M&S verification also is difficult to reconstruct. The biggest challenge to verification efforts for the IFS was the initial lack of documentation for the various modules that make up the IFS. This was in part due to lack of specificity in the contractual requirements for supporting V&V of the software; to rectify that deficiency, Raytheon subcontracted to CSC to support the development and review of Simulation

Module Design Documentation (SMDD) for each critical module within the IFS, each of which corresponded to a basic function in the overall model. There were approximately 35 SMDD reviewed by CSC, – however, some of those were simply updates written for different versions of the same model function.

It should be noted that the some of the software used in the model is the exact same software (at the source code level) used in the missile. It is just compiled for the SGI machines used by the IFS instead of the missile processor. This is true of most of the software functions in the IFS. Modules like the autopilot, tracker, TSE (Target State Estimator), and Guidance are treated in this way. Part of the functional qualification test for the missile Operational Flight Program (OFP) is to ensure that the missile algorithm is identical to the IFS algorithm when it executes. That makes for an excellent verification test for each IFS module.

The amount of work CSC did in reviewing the SMDD was spread equally across all the functions, after the module owner at Raytheon did his or her part (which in most cases was a considerable amount of work to develop the document in the first place). The total cost for the SMDD review was on the order of ½ work-years, but that does not include the cost of developing the documentation at the missile developer. It can be argued that developing that documentation (which amounts to a detailed conceptual model for each module of the IFS) is not directly part of the V&V effort, but without that documentation verification cannot proceed. One also could argue that the efforts with which JASA, CSC and SURVICE were involved were accreditation support. Model documentation is one type of evidence that Service policies and the DMSO VV&A RPG consider an important ingredient in an accreditation case. So if accreditation is the goal, the cost of developing model documentation should be included.

Based on experience with this program, there does not seem to be a strong link between the amount of work required to document a model sufficiently for accreditation and the size of the model (in terms of SLOC, for example). The amount of work required seems to depend more on whether the documentation is done by people familiar with Software Engineering concepts. It depends even more strongly on whether it’s done by people (like those at Raytheon) who are design engineers that are tasked with simulation development at the same time they develop their missile component. The more familiar they are with the system being modeled, the more efficient they can be when documenting that model, no matter how large or complex the model might be. That fact might seem self-evident, but it tends to get lost in theoretical discussions of V&V.

Summary

The AIM-9X program approach to M&S VV&A has been to integrate the V&V effort into the overall missile development process. Test firings have been used to support M&S validation reviews, missile software has been used directly in the digital simulation (thus every missile event provides support to M&S verification as well), and the accreditation and validation reviews have ensured that whatever V&V is necessary is carried out and focused on accreditation requirements.

While this integrated approach has made VV&A a cost-effective and integrated partner in overall system development, it has made it difficult to separate out resource requirements for VV&A activities. The most important function that JASA has provided to the program in its accreditation agent role is to make sure that V&V efforts are properly documented. It is these documentation efforts (and the direct costs of the accreditation and validation reviews) that can be tracked, but that does not track the total cost of V&V, only the cost of ensuring that V&V results are saved for posterity (and for use in accreditation panel reviews).

A Case Study - JSF

The goal of the Joint Strike Fighter (JSF) Program is to field an affordable next generation strike aircraft for Joint Service use. The results of M&S in a variety of engineering, design, and mission effectiveness analyses were a key factor in establishing JSF Operational Requirements. During that Requirements Development Phase, the JSF program also began a major effort to develop and execute a robust VV&A process. The best source of information on the JSF VV&A efforts and associated resource requirements is a paper by Mr. Ron Ketcham and Major Steven Bishop (USAF). The following descriptions of the JSF VV&A process and comments on resources required are liberally paraphrased from that paper.[20]

JSF M&S Tools and Data

The biggest challenge to VV&A in the JSF program is the massive size of the M&S effort the program has undertaken. The size is influenced by both the number of community standard tools JSF has chosen to use and the number and size of the M&S tools JSF is developing. JSF has developed an extensive array of both virtual and constructive M&S Tools, a key component of these tools being the Strike Warfare Collaborative Environment (SWCE). During the Requirements Phase, the SWCE contained a tool set of major models and simulations ranging from engagement level models such as ESAMS (a one threat-one aircraft missile engagement model), to the Joint Integrated Mission Model (JIMM), on up to campaign level models such as Thunder. The SWCE has grown to 28 major models and simulations; it is a government owned and managed tool set being supplied to the JSF contractor during the System Design and Development (SDD) phase.

The SDD phase has added a second significant M&S tool set which will be utilized by the contractor, Lockheed-Martin. This is the JSF Engineering and Manufacturing Collaborative Environment (EMCE). The EMCE is the Weapon System Contractor (WSC)-managed toolset focused on requirements allocation, systems engineering processes, and the design and the manufacturing of the JSF. It is estimated that the EMCE contains about 300 engineering level models. The JSF Suite of Models and Simulations (SOM&S) is the now combined tool sets of the SWCE and the EMCE.

A "complete" VV&A effort for any one of these models would require a large investment of resources. The resources that would be required to conduct an exhaustive VV&A program for the complete suite of JSF models and simulations are staggering, even for the Department of Defense. And the scope of the JSF VV&A effort is not limited to the models alone. JSF began to address the credibility of the data used by this model suite

[20] *The Application of VV&A in Promoting the Credible Employment of M&S within The Joint Strike Fighter Program*. Ronald L. Ketcham, Maj. Steven Bishop, Proceedings of the 2002 European Simulation Multiconference, Society for Computer Simulation Europe, Edited by Krzystof Amborski and Hermann Meuth, pp 705 – 709.

during the Requirements Phase, and the size of the effort has increased significantly during SDD. In terms of resources, VV&A of data will require an effort as extensive as VV&A of M&S. The amount of data to examine is huge and is contained in four highly dynamic data bases.

The JSF Accreditation Process

In order to reduce the size of the required VV&A program to manageable proportions, the JSF Requirements Directorate adopted the tailored accreditation approach developed by the SMART project and applied by JASA; it also hired JASA as its accreditation support agent. JASA developed a VV&A process for JSF that addressed the growing needs of the program to assure the credibility of its M&S. The JSF VV&A process has grown and evolved with the program, and lessons learned over time have led to improvements. For all the iterations of the Requirements Phase, JSF followed the following four-step accreditation process:

Defining the Application

The first step in the JSF accreditation process is to define in detail what questions are being addressed with M&S. Near term M&S needs that were identified by this analysis focused JSF VV&A efforts in the Requirements Phase, while mid to far term needs fed into JSF model improvement efforts designed to ensure that tools would have adequate capabilities during SDD and beyond.

Developing M&S Requirements

M&S requirements serve as acceptance criteria against which the models are assessed, and help to focus the V&V effort on the M&S elements most important to JSF. JSF evaluates four types of M&S requirements: general, functional, fidelity, and operational. General requirements address areas of broad interest to all potential users of a model (such as software quality, extent and currency of documentation, assumptions, limitations, and known errors, etc.) Functional requirements detail the functions which the model must perform in order to adequately address the questions being answered using M&S results. Fidelity requirements relate to the accuracy necessary to quantify the Measures of Effectiveness (MOE) used by JSF. And operational requirements address the environment required to use the model (computer hardware and software compatibility, etc.). Careful thought at this stage is expected to reduce the cost of V&V by focusing the work only on areas which are relevant to program needs and are supportable by the tools and data available.

Comparing M&S Capabilities against Requirements

The next step is to gather data about the models to be used and compare their capabilities with the M&S requirements (acceptance criteria). This comparison highlights areas where the model meets the assessment criteria, indicates weaknesses which reduce the credibility of model results, and reveals areas where information is insufficient to support any conclusions about credibility.

Assessing Risk and Developing an Accreditation Recommendation

The final step of the accreditation support process is to assess the risk incurred by using the model as is, and to come up with recommendations for risk mitigation (such as using another model where the model is weak, conducting additional V&V, use of alternate MOEs which the model can support, workarounds, etc.). An accreditation recommendation is then based on those analyses.

Impacts of resource requirements

JSF had a well-defined VV&A process in place during the Requirements Phase of development. However, there were shortfalls that occurred during that Phase because of the scope of the VV&A effort and the limited time available. The scope of the required effort is enormous and getting larger. It was recognized from the start that JSF would not be able to totally V&V a single model, let alone the 28 models in the SWCE. Now there are even more M&S tools with the addition of the EMCE. On top of that, there are requirements to evaluate the credibility of the associated data. JSF does not have the luxury of misusing limited V&V resources on anything but the most critical program issues. How do you manage your limited resources to address the areas which are critical to the credible application of M&S in the program? What V&V activities will mitigate the most risk to program? This requires a constant prioritization of VV&A concerns based on risks to the program. Limited VV&A resources are applied to those areas that provide the highest return in terms of overall M&S credibility.

The second biggest challenge was time. The accreditation process defined above was executed in parallel with the JSF requirements analyses. But there was little time to address M&S concerns that surfaced as a result of VV&A efforts, and in fact for the most part the results of the accreditation process were not available until after the requirements analyses were completed. In most cases JSF was left with an understanding of the limitations of the M&S and had to make a determination whether these limitations were acceptable. While improvements could be made to M&S credibility in later phases of the requirements development process, there was little direct impact of this information on the immediate results.

In general, V&V activities will only be done when they can be completed in sufficient time to impact confidence in M&S commensurate with the resources expended during a single Accreditation cycle. However, this will not, and should not be a hard and fast rule. For example, in one of the cycles during the Requirements Phase, JASA did recommend that one model not be accredited for all applications being considered. Model improvements would be needed that could not be completed in time for that analysis cycle. However, it was recognized that the long-term interests of the program required that these investments be made. As a result, the model has been improved and is capable today of meeting JSF requirements, albeit in later phases of the program.

Management of VV&A

The primary concerns of the JSF program are not with the accreditation process. The program recognizes the huge scope of the required effort, both in terms of cost and schedule. The key concerns are how to apply that process to a large and dynamic collection of M&S tools and databases in an effective and efficient manner. The JSF program is ensuring that VV&A is an integral and evolving part of M&S management within the program by instituting a Verification and Validation Action Team (VVAT) under their overall M&S management structure during the SDD phase of the program. It will be the responsibility of the VVAT to coordinate the Accreditation Process.

The scope of this effort required the program to form an organization within the program to address the VV&A problem, because it can take so much in the way of resources. This new management structure, with collaborative teams of government and industry, is explicitly formed to address JSF VV&A requirements. These teams are designed to focus V&V efforts on essential accreditation requirements to manage the credible application of M&S tools, and their associated data, in support of the verification of system performance and effectiveness. The VVAT will continuously monitor the entire VV&A process to allow the JSF program to meet its accreditation needs in the face of the large number of M&S involved.

Summary

JSF uses a large collection of models and simulations to support the various phases of system development, and the program faces a major challenge in developing the information necessary to support their accreditation within cost and schedule constraints. The total expenditures by the program on VV&A related tasking is in the multi-millions of dollars, but that has merely scratched the "credibility surface" for the total Suite of Models and Simulations. JSF has continued to evolve their Accreditation Process to meet these challenges by focusing VV&A activities on those M&S elements that are of most importance to program objectives. This allows them to allocate their VV&A resources in the most cost-effective manner. This is done in part through the formation of the VVAT, making VV&A activities an integral part of their overall program organization.

A Case Study - SMART

Published information on what it costs to conduct verification and validation for existing M&S includes a paper by Muessig[21], written at the very beginning of the efforts by JASA to support DoD weapons systems programs with cost-effective M&S credibility. The paper tracks the actual cost history of M&S V&V efforts conducted by JASA and the project which immediately preceded JASA, the OSD funded Susceptibility Model Assessment with Range Test (SMART) project. The objectives of the SMART project included demonstrating and costing out VV&A activities as applied to 5 legacy "community standard" models used within the aircraft survivability analysis community.

Categories of V&V Tasking

The paper divided V&V tasking into three general categories. The first category has to do with characterizing the general credibility of a model by asking questions such as: How is the model managed and supported? What has it been used for, by whom, and was it accredited for that use? What is its V&V history? How well is it documented? What is the quality of the software? What are the model's known assumptions, limitations and errors? Answering these questions prior to jumping into a V&V program can help to scope any V&V effort and minimize the resulting costs. It can also help a potential user screen out candidate models that will not meet his/her needs before they commit any resources to conducting further V&V on a model.

The second category of tasking has to do with a subjective determination of model "reasonableness" by reviewing a collection of objective data. The review often is conducted by SME in areas relevant to the model's functions (for example, radar systems engineers for radar functions, guidance and control specialists for missile functions, etc.). The objective data that provide grist for the review mill consist of: quality and sources of input data (data verification & validation); comparison of model outputs with intelligence data or best estimates; sensitivity analysis results; and a summary of known assumptions, limitations and errors. Although fraught with the pitfalls inherent in subjectivity, this type of expert review provides the best possible assessment of the adequacy of macro-level model results short of detailed V&V. It also can be accomplished for much lower cost than a line-by-line code verification effort, or detailed comparisons between model outputs and field test data (which can be expensive to collect).

The third category of V&V tasking is "classical" V&V: the kind of V&V everyone is afraid of. It typically consists of line-by-line verification of the code, including desk checking, software testing and comparison to design specifications, coupled with validation efforts composed of extensive comparisons of model predictions with all available sources of test data at both the detailed M&S functional level and the overall M&S output level. Because of cost, these activities are not (or at least, should not be) generally performed on an entire model without reference to an application requirement.

[21] *Cost vs. Credibility: How Much V&V Is Enough?,* Muessig, P.R., Society for Computer Simulation

Rather, the scope of detailed V&V should be tailored to each specific application and only those M&S functions essential for use in that application should be subjected to this rigorous level of activity. This level of V&V should not be conducted "for its own sake."

To further that view, the SMART project divided the M&S into "Functional Elements (FE)" so that individual functions within the model were analyzed separately, where that was possible. This is somewhat analogous to the object oriented approach taken by some more recent software developments, but unlike object oriented programming, in the "legacy" simulations that were evaluated by SMART the software describing individual functions had to be extracted from the simulation as a whole. The purpose of separating out individual functional elements (or objects) was to focus V&V efforts on those elements of the M&S which were of highest priority to user requirements, or which had the greatest impact on simulation results.

The M&S that were assessed in this project were "engineering level" simulations; they were the Advanced Low Altitude Radar Model (ALARM), Enhanced Surface-to-Air Missile Simulation (ESAMS), Radar Directed Gun Simulation (RADGUNS), the BRAWLER Air Combat model, and the Trajectory Analysis Program (TRAP). As such they were fairly high fidelity representations of systems and probably represent a moderate to high level of complexity for M&S within the Department of Defense. They were each divided into a different number of functional elements, but they all had some functions in common (such as target, environment, radar, etc.). They were divided into approximately 15 functional elements each, some more and some less.

Cost Analysis

How much did the V&V effort really cost? To answer that question, the cost of conducting V&V was tracked by a detailed Work Breakdown Structure (WBS) for VV&A tasking developed under the SMART project[22]. The LOE required to conduct V&V tasking was tracked over the life of the project by elements within the WBS, and the average costs of each V&V task and product were developed based on the individual costs for each model. These average costs are reproduced here in the table. These costs were for V&V activities conducted after the M&S were developed, in some cases by people other than the original developers, and not for V&V conducted as an integral part of development.

The results of this cost analysis were that the first category of M&S credibility tasking, to develop a characterization of the model's VV&A status and history, level of management support, and summary of assumptions, limitations and errors took on the average 18 "work-months (WM)", or about $250K in then-year average dollars. This was from start to finish: that is, from project inception to final documentation.

[22] *VV&A from A to Z*, Dr. Paul R. Muessig, David H. Hall; Dennis R. Laack; Martha L. Hoppus; Barry O'Neal, Joint Technical Coordinating Group on Aircraft Survivability (JTCG/AS), JTCG/AS 97-M-008, October 1997

WBS#	Task Name	LOE (WM)	Cost ($K)
1.1.1	Assess Current Documentation	3	
1.1.2	Assess Software Quality	3	
1.1.3	ID Assumptions, Limitations, Errors	3	
1.2.1	Decompose Model into FE's	2	
1.2.2	Define Functional Templates	1	
1.3.1	Define Configuration Baseline	1	
1.3.2	Evaluate Existing CM Procedures	2	
1.3.3	Survey Model History	3	
	TOTAL Category I Tasks	**18**	**252**
2.1.2	Prepare S/W Design Documents	2/FE	
2.2.2	Perform Sensitivity Analysis	1/FE	
	TOTAL Category II Tasks	**3/FE**	**42/FE**
3.1	Code Verification Tasks	2/FE	
3.2	Validation Analysis Tasks	6/FE	
	TOTAL Category III Tasks	**8/FE**	**112/FE**
	GRAND TOTALS	**18+11/FE**	**252+154/FE**

Average V&V Costs Experienced by SMART Project

The second category of V&V tasking, that of developing information required to support an expert review of the model's credibility, including data V&V, sensitivity analyses, and comparisons with "authoritative sources", took an average of 3 "work-months" per FE, or approximately $42K per FE (at the time). This included data reviews, sensitivity analyses and documenting the algorithms used to describe the functional element and their sources. This is equivalent to developing a Software Design Document (SDD) for each FE. This resource estimate did not include actually conducting an expert review of the model, only the cost of developing the information required to make a review effective.

The third category of V&V tasking, detailed verification and validation efforts, took an average of 2 work-months for verification of each FE and 6 work-months for validation (for a total of approximately $112K per FE). The validation tasking did not include the costs associated with collection of test data for comparison with the model. The validation costs only included those costs associated with analyzing the test data, running the model for the same conditions as the test data, making the comparison and explaining differences (and documentation). Not all functional elements were verified and validated by the SMART project; only one simulation (ALARM) was entirely verified (all FE), and none had validation activity for all functional elements. Even with a project devoted specifically to demonstrating the V&V process for a set of 5 simulations, this third category of V&V activity proved too costly to conduct in an exhaustive manner. It's no wonder that many program managers think that "V&V costs too much and takes too long."

It should also be pointed out that the functional element approach was not necessarily the least cost method for verification tasking. During the SMART project the principal independent verification agent, ENTEK, Inc. pointed out that the most natural way to examine code is to follow the inherent structure of the code itself; i.e., the call hierarchy[23]. However, for the legacy M&S examined during SMART, the code structures did not align themselves naturally with the functional elements of the systems being modeled. So while the functional element approach provided for efficiency in validation, the extra effort involved in "re-organizing" verification efforts around functions within the code resulted in verification being less efficient than a traditional approach. However, the cost benefits accrued by the validation activity more than offset the relatively small increase in verification costs.

There are some interesting characteristics about this approach to V&V that should comfort those concerned about runaway V&V costs. For example, the first type of V&V tasking, whose products relate to the credibility of the model as a whole, has a fixed cost (i.e., about 1 ½ work-years for a complete documentation of the results), while the V&V tasks described in the other categories have a variable cost related to the number of specific M&S functions requiring credibility to support a particular application. Since the total cost of all V&V tasks for two FE exceeds the cost of the entire first look at the model's credibility, it is clear that V&V costs are dominated by the number of FE that are important to a particular application. This suggests that significant attention must be paid to defining the smallest possible set of model functions that are critical to the model's use for a given application.

[23] Pilot Verification Study Report, ENTEK/ABQ-94-0106-TR, February 16, 1994

Analysis Of State Of The Art And Program Case Histories

Common Experiences of Case Study Examples

There are a number of common elements that cut across the case studies that were presented above, in terms of their experiences with determining V&V resource requirements. These have to do with the process of identifying required V&V activities, the difficulty of separating V&V activities from M&S development, and the use of risk as a determinate of V&V requirements.

Developing Accreditation (and M&S) Requirements in the Design of a V&V Program

Due to constrained resources, none of the programs examined in the case studies were able to conduct an exhaustive V&V program (nor did they desire to do so). The greatest challenge to all of these programs was to determine what is "Good Enough" testing of the software and good enough correlation with test data to constitute sufficient M&S validation. They required an objective way to narrow the scope of required V&V activities based on the requirements of their application.

Requirements Based on Risk

All of the approaches to VV&A tasking we have discovered, and the one cost estimating tool described in this paper base the required scope of V&V activities on an assessment of risk. In all of these approaches, risk is defined as probability of occurrence multiplied by the impact of adverse events, but none of the approaches makes an attempt to calculate risk directly. Both the JASA and UK approaches make use of the risk matrices developed by the safety community to provide a subjective, but structured assessment of risk and the resulting V&V activities that are required. The Managed Investment Approach also addresses risk, although in a slightly different manner. And the Cost Estimating Tool uses Risk and Uncertainty factors in the estimate of V&V resource requirements.

The case studies almost uniformly apply a risk-based approach to determining V&V requirements and tasking. This is partly because the case studies involve programs which executed the approaches described elsewhere in this paper, but they are representative of the type of thinking that typically goes into the V&V resource problem. Everyone asks the question, "How wrong can my M&S be and still give me the right answer to my question?" where the "right answer" is one that results in little risk to the program.

Paucity of Detailed Historical Data on V&V Resource Requirements

There is very little historical cost data available to describe V&V cost experience. The Hicks study pointed out that not only is there little V&V cost data, but there is little M&S development cost data of any fidelity or credibility within the DOD. The only detailed cost data we found was from the SMART project, which developed and used a detailed WBS to track project V&V costs over a number of M&S.

This is in part due to the fact that most DOD programs (with the possible exception of software development programs) do not track M&S costs very well. Development of the M&S is not their primary goal, so they do not track the costs of the M&S separately from the costs of the overall system development. M&S development, and application cost data are not readily available within acquisition programs.

There is a lack of management visibility into program expenditures for M&S activities in general, in part because standard cost accounting procedures do not provide for segregation, reporting or tracking of M&S costs. Another complicating factor is that M&S activities often are not listed as deliverable items in contracts, meaning that the contractor is not under any obligation to report the expenses associated with M&S activities even if they could do so. And programs themselves are not required to track M&S expenditures, so they don't track them.

The lack of historical V&V cost data is also in part due to the practice of integrating V&V into software development, which is generally a good thing. But while this integrated approach has made VV&A a cost-effective and integrated partner in overall system development, it has made it difficult to separate out resource requirements for VV&A activities.

Another factor complicating the task of tracking VV&A costs within programs is the political nature of M&S accreditation. There are often a number of participants in Expert Reviews, for example, who are participating more because of the agency they come from than for their particular technical expertise. In order to get "buy-in" to the accreditation decision, these participants need to be part of the accreditation review process. This is not necessarily an indictment of the political nature of program management, but it does make it difficult to predict what the required reviews will cost in general if you don't know well ahead of time who you need to invite.

One of the most important functions that a V&V or Accreditation agent can provide to a program is to make sure that V&V efforts are properly documented. It is these documentation efforts (and the direct costs of accreditation and validation reviews) that can and have been tracked by programs, but that does not track the total cost of V&V, only the cost of documenting it.

Constrained costs

No matter what V&V resource requirements are determined to be, based on risk or other factors, the common experience of all the programs examined in the case studies were that they were not allowed to spend that much. For example, in the GBR HWIL Testbed development effort the resources available to the conducting agencies constrained what V&V activities were undertaken, what V&V products were generated, and the degree to which all the V&V needs of this program were met. Consequently, the GBR PM needed to execute a plan of V&V activity which was responsive to the needs of the BMDO accreditation authority and other potential accreditation agencies, and was practical to implement. The level of funding available established the LOE for the V&V program and directly impacted the selection of V&V activities and their application to GBR HWIL units-under-test.

This was also true of every other program we examined (and it should come as no surprise – nobody can afford an exhaustive V&V program, and apparently nobody can afford the one they need, either). Even for the JWARS program, which was a large software development (as opposed to the missile, aircraft, and other system acquisition programs for which the development of M&S was a by-product), the V&V effort was resource constrained. The total JWARS V&V program could not exceed $2.5M; that number was decided upon up front, before any requirements analysis could be performed.

Even though all of the approaches to V&V planning were described as determining the level of M&S credibility required to support an accreditation decision by the program sponsor, all of the practical examples of V&V programs we have included were faced with a "fixed resources - variable benefits" problem: they had to determine as best they could what V&V activities would provide the maximum benefit for the constrained resources available. Because they were not allowed to exercise a program that would provide the full credibility required, that meant they had to fall back on identifying the risk associated with using the M&S anyway.

Documentation Costs

The common experience among the case study examples was that during development none of the simulations were documented adequately to support the required VV&A activities. All of the programs had to supplement the M&S documentation, at their own expense. Based on the experiences of a couple of these programs, and the experience of JASA on other programs as well, there does not seem to be a strong link between the amount of work required to document a model sufficiently (to support VV&A) and the size of the model (in terms of lines of code, for example). The amount of work required seems to depend more on whether the documentation is done by people familiar with Software Engineering concepts. It depends even more strongly on whether it's done by people (like those at Raytheon for the AIM-9X program) who are design engineers that are tasked with simulation development at the same time they develop their missile component. The more familiar they are with the system being modeled, the more efficient they can be when documenting that model, no matter how large or complex the model might be. So independent of the cost of doing V&V, the cost of documenting the M&S to support V&V appears to be relatively fixed.

Schedule Impact of V&V

Often a program has already decided which M&S they are going to use and has committed to use those M&S before any VV&A effort starts. While this is common practice, especially for programs which are using M&S that represent their own system, it does generally result in analysis being done with the M&S in parallel with the VV&A effort. This actually can increase the risk to the program, both in terms of technical risk and schedule risk. In terms of technical risk, it may be that M&S results are used to make program decisions before V&V results are available, which may mean that those decisions could end up being wrong if the M&S have errors that were undetected prior to the V&V effort. In terms of schedule risk, the analysis may have to be done over with an updated model version because the V&V found and fixed errors in the version the analysts were using. Or it ends up as in the JSF case with the accreditation assessment really being an assessment of the risk of using the results you've already generated.

What this really means is that in many practical cases, for system acquisition programs we find that the schedule for V&V is not in synch with the schedule for using M&S results. This is true not only for development but also for operational testing. While AIM-9X is a sterling example of the use of M&S in system development, even for that program there was often insufficient time to provide for what was considered to be adequate V&V prior to using the M&S. And the use of OT&E test results to support M&S validation can only be accomplished after the shot occurs, which makes analyses conducted prior to the test less certain than afterward. But perhaps that is just the way life is for use of M&S to support operational testing.

Cost Estimating Relationships

The ultimate goal of this paper was to examine the state of the art and make recommendations for estimating the resource requirements for M&S VV&A activities. The closest we are likely to come to recommendations is to identify those factors that should be included in cost estimating relationships. The CET, by Robert O. Lewis appears to be the current state of the art for estimating VV&A resource requirements; it is a parametric cost model that considers the inherent complexity of each software package, the effects of leveraging and risk and uncertainty, and then uses size of the software as the major variable, except for federations. Federations use a percentage of the development cost as the primary variable.

Primary factors that must go into the VV&A cost estimate include: size (LOC, SLOC) and programming languages, complexity, intended use of the product, risk and uncertainty, unique characteristics of the M&S, amount of reuse, any previous VV&A

history, availability of appropriate tools, status of formal documentation, and adequacy of the documentation. Also included are costs of SME, software tools, support software and hardware, communications and networking, travel and TDY, and any other extraordinary costs that may occur. We also must include somehow the level of expertise of the practitioners involved in the V&V tasking; their experience level drives not only the resource requirements but the quality of the final product. And the CET does not currently cover the cost of simulation validation, especially if the costs of test data collection are to be included in validation cost estimates.

Leveraging

Consider two users of the same model who wish to accredit it for different applications. Assume further that both users have decided to summarize what's known about the model's credibility, and that both require detailed V&V for the same 5 high importance functions within the model.

If they both work independently, and have no knowledge of each other's efforts, they will each spend (using the cost data accumulated by the SMART project) almost a half-million dollars on V&V. At the end of all that, each user will have completed the same set of V&V tasks. Several important questions arise: How did each user interpret the same V&V task? Based on the different possible interpretations of each V&V task, how different do the resultant V&V reports look? Based on how different the V&V reports might look, how can future users benefit from two reports that interpret V&V differently, that say different things about the same model, and that report those things in different ways? Could the two efforts have worked together to reduce the total V&V cost burden while still accomplishing their individual accreditation objectives? How can future users benefit from the work these two independent efforts have done?

If each effort had known of the other, they could have drawn up a consolidated list of V&V tasks by coordinating their requirements. In addition, a common understanding of V&V tasks could have been developed, as well as an understanding of how these tasks related to the previously defined fidelity requirements. Such a review could have led to a common V&V reporting template that would have served the needs of both applications without duplication of effort. By making this common V&V report available to other users of the same model, V&V requirements in support of other applications could be reduced substantially, by forming the nucleus of a body of evidence supporting the model's credibility, a body to which other users with other applications could have contributed, in turn.

Now let's assume that a third user is required to accredit the same simulation, but he determines that 5 <u>different</u> functions are important to his problem. If he develops the V&V information required for those functions, and reports them in the same format that the other two users developed, then V&V results are now available for 10 of the primary functions in the model (or about 2/3 of the model, if the model is like the ones documented by SMART). By incrementally adding to the body of knowledge about the model's credibility, we as a community of users can eventually provide information on the credibility of the entire simulation, one function at a time. And no one user has to pay for it all.

Summary

Risk is a primary driver of VV&A resource requirements. We have described several approaches to linking risk with V&V requirements, and have shown how several real world programs have executed those approaches. However, we have also shown that in the real world programs are constrained by cost and schedule to such an extent that even V&V programs that are tailored to their specific requirements are not executable. Schedule demands also militate against effective use of V&V results in time to impact program decisions that are based, in part, on M&S results.

There is very little historical VV&A cost data available, largely because programs do not track those costs in any detail. And it is difficult to separate VV&A costs from software (or system) development costs, because of the laudable trend toward integrating V&V activities into software development. There is no universally accepted definition within the community for how to separate those costs – what tasks are included in VV&A, and what tasks are not.

Recommendations

The discussion and analysis above have highlighted a number of areas that warrant further research and evaluation, summarized below.

VV&A Tasking Definitions and WBS

If we are to identify VV&A resource requirements, whether costs, manpower or schedule, we must have standard definitions of which activities that are conducted during simulation development are VV&A activities, and not development or other activities. This can best be done by adopting a generally accepted standard work-breakdown structure for VV&A. If we had a standard VV&A WBS that was used by the entire community for tracking costs (and for task planning), then those costs would be consistently collected and uniformly analyzed. Significant care will need to be taken in deriving these definitions and the accompanying WBS; tasks such as test range data collection, analysis, documentation and comparison with M&S results can support simulation validation, system development, system testing, or all of the above.

Cost Data Collection

There is a dire need for more real-world M&S VV&A cost data. This can be supported by the previous recommendation: once we have common accepted definitions of VV&A tasking, and an associated WBS, we can begin to collect meaningful data on the resources expended in accomplishing those tasks. In order to collect such cost data, we will need to convince system acquisition and software development programs of the benefits of such a standardized approach to VV&A task planning and tracking. By following this approach, with such a WBS, programs can be assured that they are adopting a cost-effective VV&A process as well as providing the necessary data to improve our abilities to predict the resources required to execute the process. Side benefits to those programs include adequate descriptions of required deliverables in contract negotiations, thus minimizing the inevitable contract re-negotiations and misunderstandings between contractors and government sponsors when it comes to VV&A tasking.

Data Pedigree for Cost Estimating Relationships (CER)

If we can collect data from the tasks identified by a standard WBS, we can start to develop the required validation and calibration data for cost estimating relationships. These data can be used to support and "validate" the algorithms and CER in the CET, for example. The CET provides an excellent framework to develop a valuable VV&A resource estimating tool, but it needs a pedigree for the supporting data used to derive the CER, and validation data to demonstrate its utility. If we can develop pedigreed data for CET inputs and validation data to show its credibility, we will be providing a stronger link between the academic community which provides the theory behind estimating VV&A resources and the practical application of those theories.

Acronym List

Acronym	Meaning
A	Accreditation
ABM	Anti Ballistic Missile
AFOTEC	Air Force Operational Test and Evaluation Center
AFSAA	Air Force Studies and Analysis Agency
AIM	Air Intercept Missile
ALARM	Advanced Low Altitude Radar Model
ALSP	Aggregate Level Simulation Protocol
ATEC	Army Test and Evaluation Command
AZ	Azimuth
BM/C3	Battle Management and Command, Control and Communications
BMDO	Ballistic Missile Defense Organization
CAA	Center for Army Analysis
CAETI	Computer Aided Education and Training Initiative
CCB	Configuration Control Board
CET	Cost Estimating Tool
CER	Cost Estimating Relationship
CM	Configuration Management
CMM	Capability Maturity Model
CMMS	Conceptual Model of the Mission Space
COMOPTEVFOR	Commander, Operational Test and Evaluation Force
COTS	Commercial Off The Shelf
CSC	Computer Sciences Corporation
DA PAM	Department of the Army Pamphlet
DARPA	Defense Advanced Research Projects Agency
DEM/VAL	Demonstration/Validation
DERA	Defense Evaluation and Research Agency
DIS	Distributed Interactive Simulation
DMSO	Defense Modeling and Simulation Office
DoD	Department of Defense
DODI	Department of Defense Instruction
DOT&E	Director, Operational Test and Evaluation
DT&E	Development Testing and Evaluation
ECM	Electronic Countermeasures
EL	Elevation
EMCE	Engineering and Manufacturing Collaborative Environment
ESAMS	Enhanced Surface to Air Missile Simulation
FE	Functional Element
FEDEP	Federation Development and Execution Process
FFRDC	Federally Funded Research and Development Center
FY	Fiscal Year
GBR-P	Ground Based Radar - Prototype
GMD	Ground-Based Midcourse Defense
GMD JPO	Ground-Based Midcourse Defense Joint Program Office
GOTS	Government Off The Shelf

HLA	High Level Architecture
HW	Hardware
HWIL	Hardware in the Loop
I/O	Input/Output
ID	Identify
IEEE	Institute of Electrical and Electronics Engineers
IFS	Integrated Flight Simulation
IMC	Innovative Management Concepts
IMU	Inertial Measurement Unit
IPT	Integrated Product Team
ISO	International Standards Organization
ISTC	Integrated System Test Capability
ITOP	International Test Operating Procedures
IV&V	Independent Verification and Validation
JAD	Joint Application Design
JAMIP	Joint Analytic Model Improvement Program
JASA	Joint Accreditation Support Activity
JIMM	Joint Integrated Mission Model
JSEM	Joint Services Endgame Model
JSF	Joint Strike Fighter
JST	JWARS Study Team
JWARS	Joint Warfare System
KMCC	Kwajalein Mission Control Center
LFT&E	Live Fire Test and Evaluation
LOC	Lines of Code
LOE	Level of Effort
M&S	Models and Simulations
MCCDC	Marine Corps Combat Development Command
MDA	Missile Defense Agency
MDAP	Major Defense Acquisition Program
MDSE	Missile Defense System Exerciser
MOE	Measure of Effectiveness
NAWCWD	Naval Air Warfare Center Weapons Division
NIST	National Institute of Standards and Technology
NMD JPO	National Missile Defense Joint Program Office
OD PA&E	Office of the Director, Program Analysis and Evaluation
ODC	Other Direct Costs
OFP	Operational Flight Program
ORD	Operational Requirements Document
OSD	Office of the Secretary of Defense
OSD/SAC	OSD Studies and Analysis Center
OT&E	Operational Test and Evaluation
OTA	Operational Test Activity
PAC-3	Patriot Advanced Capability III
PCIL	Prime Consolidated Integration Laboratory
PIPT	Program Integrated Product Team
PM	Program Manager
R&U	Risk and Uncertainty
RADGUNS	Radar Directed Gun Simulation
ROI	Return on Investment

RPG	Recommended Practices Guide
S&T	Science and Technology
SAC	Studies and Analysis Center
SCTV	Separation/Control Test Vehicle
SDD	System Design and Development
SDD	Software Design Document
SDP	Software Development Process
SEI	Software Engineering Institute
SEPG	Software Engineering Process Group
SETA	Systems Engineering and Technical Assistance
SGI	Silicon Graphics
SLOC	Source Lines Of Code
SMART	Susceptibility Model Assessment with Range Test
SMDD	Simulation Module Design Documentation
SME	Subject Matter Expert
SOM&S	Suite of Models and Simulations
SW	Software
SWCE	Strike Warfare Collaborative Environment
SWIL	Software in the Loop
T&E	Test and Evaluation
TAP	Technical Assessment Procedure
TDY	Temporary Duty
TILV	Target Interaction, Lethality and Vulnerability
TM	Telemetry
TRAP	Trajectory Analysis Program
TSE	Target State Estimator
TWG	Technical Working Group
U.K.	United Kingdom
U.S.	United States
UML	Unified Modeling Language
USAKA	US Army Kwajalein Atoll
UUT	Unit Under Test
V&V	Verification and Validation
VV&A	Verification, Validation and Accreditation
VVAT	Verification and Validation Action Team
VVT&E	Verification, Validation, Test and Evaluation
WBS	Work Breakdown Structure
WIPT	Working Integrated Product Team
WM	Work-Month
WSC	Weapon System Contractor
XBR	X-Band Radar